Charles University

Faculty of Science

Study programme: Biology
Branch of study: Biology

Eliška Mikesková

Regulace genové exprese v aktinobakteriích

Regulation of gene expression in actinobacteria

Bachelor’s thesis

Supervisor: Mgr. Jarmila Hnilicová, Ph.D.

Prague, 2025

Declaration:

I hereby declare that I have developed this Thesis independently, using only the listed literature and resources. This Thesis or a substantial part of it has not been previously used to gain this or a different academic title.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 28.04.2025

Podpis

Acknowledgements

I would like to thank my supervisor Mgr. Jarmila Hnilicová, Ph.D. for her guidance during the writing of this thesis. Apart from sharing valuable insights, she welcomed me into her research team of the Laboratory of Regulatory RNAs, which I am also grateful for.

Abstract

Regulating gene expression is greatly important for the survival of bacteria. Many of these regulations occur on a transcriptional level by regulating the transcription machinery. A necessary component of the transcription machinery is the RNA polymerase (RNAP) holoenzyme, formed by the RNA polymerase core binding a sigma (σ) factor. In this Thesis, I compare transcription regulation in various species of *Actinobacteria*. Transcription regulation varies across the *Actinobacteria* phylum and thus these regulations are divided by genus. The overview further focuses on regulatory sRNAs and different types of σ factors involved in transcription regulation. The mainly discussed σ factors are the primary σ^A and primary-like σ^B. The transcription factors RbpA and CarD present in *Mycobacteria* bind the RNAP holoenzyme and stabilise the open promoter complex (RPo). RbpA is described both in *Mycobacteria* and *Streptomyces*. Furthermore, in *Mycobacteria* the protein HelD is important for RNAP recycling while MoaB2 inhibits transcription of housekeeping genes by sequestering σ^A. *Bifidobacteria* utilise WhiB-like family proteins in response to stresses. *Actinobacteria* also use regulatory sRNAs such as Ms1 RNA found in *Mycobacteria* and *Streptomyces* or CoRP RNA specific to *Corynebacteria*. Overall, this Thesis characterises how various regulators are used by *Actinobacteria* to regulate transcription.

Key words: RNA polymerase, sigma A, Bifidobacteria, transcription, sRNA, Ms1

Abstrakt

Regulování genové exprese je velmi důležité pro přežití bakterií. Mnohé z těchto nezbytných regulací se uskutečňují na úrovni transkripce skrze regulování transkripčního aparátu. Nutnou součástí transkripčního aparátu je holoenzym RNA polymerázy (RNAP), který vzniká, když jádro RNA polymerázy naváže sigma (σ) faktor. V této práci porovnávám regulace transkripce u různých druhů z kmenu *Actinobacteria*. Regulace transkripce se značně liší v rámci kmenu *Actinobacteria*, tudíž byly regulace rozděleny podle rodů. Tento přehled se dále zaměřuje na regulační sRNA a různé typy σ faktorů, které se podílejí na regulaci transkripce. Hlavní zaměření je na alternativní σ faktor, σ^B, který je blízký primárnímu σ faktoru, σ^A. Transkripční faktory RbpA a CarD v mykobakteriích váží holoenzym RNAP a stabilizují otevřený komplex promotoru. RbpA je popsán u mykobakterií i streptomycet. U mykobakterií dále protein HelD recykluje RNAP a protein MoaB2 inhibuje transkripci údržbových genů skrze vyvazování σ^A. Bifidobakterie využívají proteiny rodiny WhiB-like při stresu. Aktinobakterie také využívají regulační sRNA jako je Ms1 RNA u mykobakterií a streptomycet nebo CoRP RNA u corynebakterií. Souhrnně řečeno, tato práce popisuje, jak aktinobakterie využívají různé regulátory pro regulaci transkripce.

Klíčová slova: RNA polymeráza, sigma A, bifidobakterie, transkripce, sRNA, Ms1

List of frequently used abbreviations

DNA - deoxyribonucleic acid

RNA - ribonucleic acid

RNAP - RNA polymerase

NCR - non-conserved region

NTD – N-terminal domain

CTD – C-terminal domain

RbpA - RNA polymerase-binding protein A

CarD-RID - CarD-RNAP interacting domain

CoRP RNA - Corynebacterium RNAP-binding RNA

RIP-seq – RNA immunoprecipitation sequencing

RPo – open promoter complex

RPc – closed promoter complex

EC – elongation complex

ECF – extracytoplasmic function group

1D – one-dimensional

sRNA – small RNA

rRNA – ribosomal RNA

NTPs – nucleoside triphosphates

bp - base pairs

nt – nucleotides

List of used abbreviations of amino acids

A - alanine

C - cysteine

E – glutamic acid

G - glycine

I - isoleucine

L - leucine

R - arginine

S - serine

V - valine

W - tryptophan

List of used abbreviations of nucleotides

A - adenine

T - thymine

G - guanine

C - cytosine

U - uridine

Table of contents

Abstract iv

List of frequently used abbreviations vi

1. Introduction 1

2. Transcription 1

2.1. Bacterial RNA polymerase 1

2.2. Initiation and elongation 3

2.3. Termination 3

2.4. Sigma factors 4

3. Bacterial growth curve 5

4. Regulatory proteins binding RNA polymerase 6

5. Regulatory RNAs binding RNA polymerase 7

5.1. Regulatory RNAs binding RNA polymerase specific to *Actinobacteria* 7

6. *Actinobacteria* and regulation of gene expression 7

6.1. Mycobacterium 9

6.1.1 RbpA 9

6.1.2 CarD 12

6.1.3 HelD 16

6.1.4 MoaB2 18

6.1.5 Ms1 RNA 19

6.2. Streptomyces 21

6.2.1 Ms1 RNA 21

6.2.2 RbpA 22

6.3. Corynebacterium 23

6.3.1 CoRP RNA 23

6.4. Bifidobacterium 25

6.4.1 Promoters and σ factors 25

6.4.2 WhiB-like family proteins 26

7. Conclusion 28

8. References 30

1. Introduction

As described by the Central dogma of molecular biology, transcription is a key process which allows information encoded in the genome to be expressed in the form of proteins or non-coding RNA. During transcription, genes are rewritten into RNA by DNA dependent RNA polymerase. The noncoding DNA strand serves as a template for RNA polymerase to synthesise the complementary, coding RNA strand (Crick, 1970). It is a complex process involving many other factors apart from RNA polymerase, providing the necessary space for appropriate gene expression regulation. In bacteria, gene expression is often regulated on a transcriptional level. Transcription regulation can occur by regulating the bacterial transcription machinery. Commonly, the activity of RNA polymerase (RNAP) and σ factors is modulated.

The domain *Bacteria* is rich and diverse, consisting of various phyla such as *Proteobacteria* and *Actinobacteria*. Research on *Actinobacteria* greatly focuses on their production of unique secondary metabolites. Importantly, the dangerous pathogen, *Mycobacterium tuberculosis*, belongs to the phylum *Actinobacteria*. The well-known *Escherichia coli* belongs to the phylum *Proteobacteria*. *E. coli* has served as a model for understanding most cellular regulations, including transcription regulation of RNAP (reviewed in Ruiz & Silhavy, 2022). Actinobacterial RNAP and proteobacterial RNAP experienced divergent evolution and thus differ in many ways (Lane & Darst, 2010). Their divergent evolution thus encourages for further research of gene expression regulation to focus on species of *Actinobacteria*. Despite not being researched as closely as *E. coli,* many regulators present in *Actinobacteria* have been characterised. Some of them are specific only to *Actinobacteria*.

The aim of this Thesis is to provide an overview of transcription regulation in various species of *Actinobacteria*. It will focus on regulatory RNAs and different types of σ factors involved in transcription regulation. Among the discussed regulatory RNAs will be Ms1 RNA present in *Mycobacteria* and the recently discovered CoRP RNA present in *Corynebacteria*. The mainly discussed σ factors involved in transcription regulation will be the primary σ^A and primary-like σ^B.

2. Transcription

2.1. Bacterial RNA polymerase

The multisubunit enzyme necessary for transcription in bacteria is the DNA-dependent RNA polymerase (RNAP). RNAP is found in the form of a core enzyme or a holoenzyme. The holoenzyme is formed when the sigma subunit (σ factor) binds to the RNAP core. In the form of a holoenzyme, the RNAP is able to associate with a promoter and initiate transcription (Burgess et al., 1969). In *E. coli* the RNAP core typically consists of two alpha, two beta, and one omega subunit (αI, αII, β, β´, ω). The essential subunits of the RNAP core are αI, αII, β and β´. Unlike eukaryotes, there is only one type of bacterial RNAP, and it is essential during the entire transcription process. Nonetheless, subunits of the bacterial RNAP and subunits of the various eukaryotic RNAPs are considered homologs by structure and by function (Minakhin et al., 2001; Zhang et al., 1999).

The primary functions of each of the subunits vary. The main function of the dimerising α subunits is to assemble the RNAP (specifically the β and β′ subunits) and regulate transcription. Dimerization of the two α subunits occurs through their N-terminal domains. Each of the subunits interacts with one of the catalytic subunits, meaning αI interacts with β and αII interacts with β′. An example of the function of the C-terminal domain is its interaction with the upstream element (UP element) present in some promoters (Blatter et al., 1994). The β subunits are catalytic, meaning mainly responsible for synthesis of the polyribonucleotide strand. The β subunits consist of multiple domains. The presence of the ω subunit is associated with a higher proportion of correctly assembled and stable RNAP (Mukherjee & Chatterji, 1997). It is in contact with the β′ subunit and is a key part of assembling the β′ with the β, aI, aII assembly intermediate, thus acting as a chaperone. Having assembled the RNAP core, the σ subunit (σ factor) can take part in forming the holoenzyme and performing its functions (Burgess et al., 1969). The σ subunit which allows for promoter-directed initiation of transcription is responsible for recognising specific promoter sequences, positioning of the RNAP at its target promoter and melting the double helix DNA at the transcript start site. (reviewed in Browning & Busby, 2004; Minakhin et al., 2001; Zhang et al., 1999).

As researched on *Thermus aquaticus,* the RNAP core has a crab claw-like structure (Zhang et al., 1999). A model of the structure of the RNAP core can be seen in Fig 1. The two parts of the claw are primarily formed by the catalytic subunits and between them is a large channel, the primary channel, with the active centre where a Mg^{2+} ion is located. The primary channel is where the DNA template and the DNA-RNA hybrid are during transcription. RNAP also has a different channel, the secondary channel. Besides having access to the active site, it allows substrate NTPs to enter (Zhang et al., 1999). RNAP also has the RNA exit channel through which the nascent RNA transcript exits (reviewed in Uptain et al., 1997).

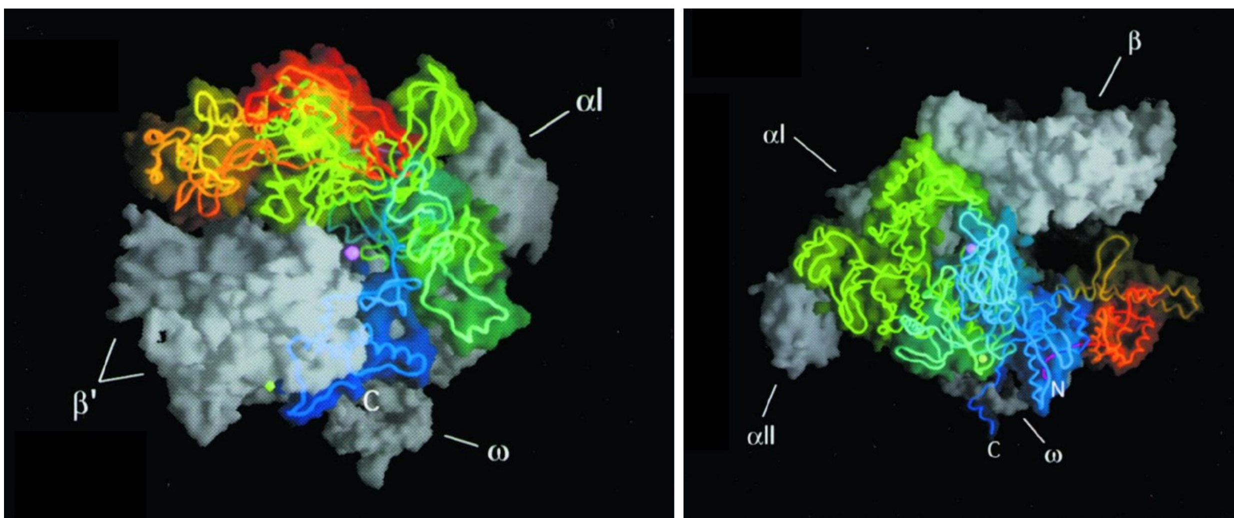

Figure 1. Structure of RNAP core from *T. aquaticus.* The figure shows two views of a 3D model of the RNAP core structure obtained based on a crystal structure. The mentioned subunits of the RNAP core are labelled. A magenta sphere represents the Mg^{2+} in the active centre. The letters "C" and "N" label the C- and N-terminus of the β subunit (left image) and the β′ subunit (right image). The backbone of the polypeptide chain of the β subunit (left image) and of the β′ subunit (right image) are shown in colour. Adapted and modified from (Zhang et al., 1999).

2.2. Initiation and elongation

Transcription can be divided into three main stages, the first being transcription initiation which results in the formation of an open complex. As mentioned before, the RNAP core must interact with a σ subunit to form a holoenzyme. The holoenzyme can then interact with a promoter sequence. The promoter sequence is a specific region in the genome upstream of a gene to be transcribed. It can consist of five elements: the -10 element, the -35 element, the extended -10 element, the discriminator and the upstream element (UP element). The primary elements are the -10 element and the -35 element, and they are located 10 bp and 35 bp upstream of the transcription start site (TSS) labelled +1 (Dickson et al., 1975). The holoenzyme interacts with the elements of the promoter mainly via different domains of the σ subunit, the only exception being the UP element as mentioned previously in chapter 2.1 RNA polymerase. The elements are involved in what is called the promoter recognition (reviewed in Browning & Busby, 2004).

Following the promoter recognition, an isomerisation occurs, creating a so-called bubble within the genome in the region of -10 to +2 (Tomsic et al., 2001). During isomerisation, the non-template DNA strand becomes bound by domain 2 of the σ subunit and the template strand moves into the active site of RNAP. The isomerisation leads to the forming of the open promoter complex (RPo) from the closed promoter complex (RPc) (reviewed in Browning & Busby, 2004). After forming the RPo, it synthesises short RNA transcripts about 2 - 15 nucleotides in length. These short transcripts are released as part of abortive transcription. During abortive transcription, the RNAP stays bound to the promoter (Carpousis & Gralla, 1980). The synthesis of long nascent RNA transcripts is associated with promoter escape. During promoter escape the RNAP holoenzyme loses contact with the promoter and moves downstream of the promoter. Importantly, the nascent RNA is held more tightly by the elongation complex (EC). Whether the σ factor is released from the RNAP holoenzyme during promoter escape is unclear. However, the σ factor is certainly displaced from the -35 and -10 promoter elements (reviewed in Hsu, 2002). The EC then synthesises the nascent RNA transcript. The addition of NTPs complementary to the DNA template occurs at the active centre of RNAP. This forms an RNA-DNA hybrid and as the synthesis continues, the synthesised RNA strand moves and exits through the RNA exit channel (reviewed in Uptain et al., 1997).

2.3. Termination

As reviewed by Santangelo & Artsimovitch (2011), termination of transcription can be categorised into intrinsic termination and factor-dependent termination. Intrinsic termination occurs via an intrinsic site and does not require any additional factors. The EC recognises a nucleic acid signal, the intrinsic terminator signal, and dissociates from the nascent RNA and the DNA template. Intrinsic termination can be adjusted by accessory proteins, but unlike factor-dependent termination, it does not require factors to recognise the termination signal. Factor-dependent termination requires regulatory proteins. One of the most well-known is Rho-dependent termination. Intrinsic termination is also known as Rho-independent termination (reviewed in Santangelo & Artsimovitch, 2011).

A well characterised intrinsic terminator is commonly found in *E. coli* (Carafa et al., 1990). Due to the intrinsic termination signal, the transcribed RNA forms a stable hairpin and has a series of uridine

nucleotides (the U stretch). Due to the hairpin and low stability between the uridine and adenine residues of the RNA-DNA hybrid, the RNAP pauses and then dissociates, releasing the nascent RNA (Farnham & Platt, 1980; Martin & Tinoco, 1980).

Rho, the termination factor first isolated by Jefferey Roberts from *E. coli*, is required for Rho-dependent termination in bacteria (Roberts, 1969). The nascent RNA has rut sites (Rho utilisation sites) which Rho binds to via its primary RNA-binding surface (Thomsen et al., 2016). Research by Song *et al.,* (2022) focused on the exact mechanisms of this termination since opposing models have been proposed. They observed that one terminator could be terminated by Rho in three different ways. To summarise, they found that Rho can both bind to a nascent RNA rut site and then catch-up to the EC pausing at a termination site (catch-up model) or it can bind to the EC and wait for the nascent RNA rut site to emerge from the RNA exit channel (stand-by model). They further observed 1D (one-dimensional) recycling, where the nascent RNA is released and RNAP diffuses along the DNA template until it reinitiates transcription or falls off. However, they also observed one-step disassembly of EC, where the RNA and RNAP dissociate from the DNA template at the same time. From Rho-dependent terminators, one-step disassembly of EC was observed more frequently than 1D recycling. RNA shearing, pulling the nascent RNA from the RNA-DNA hybrid, always followed the catch-up mechanism. Apart from RNA shearing, RNA can also be released by RNAP displacement. According to their results, the catch-up model can be followed by RNA shearing or RNAP displacement. RNA shearing would then lead to 1D recycling and RNAP displacement would then lead to disassembly of EC. The stand-by model would be followed by RNAP displacement which would then lead to disassembly of EC (Song et al., 2022).

2.4. Sigma factors

As mentioned previously in chapter 2.1.1., σ factors direct the RNAP core to transcription start sites to initiate promoter-specific transcription (Dombroski et al., 1993). It also allows the transcription bubble to form by initiating the separation of the dsDNA template. The primary σ factor of *E. coli*, σ^{70}, is involved in transcribing thousands of genes necessary for the cell's growth. These genes are called housekeeping genes. σ^{54} is not structurally related to σ^{70} and is involved in transcribing genes according to signals from the environment. In general, σ factors have structured helical domains which are connected by flexible linkers (reviewed in Feklístov et al., 2014).

According to structural homology, σ factors belong either to the σ^{70}-family or the σ^{54}-family (Merrick, 1993). Most σ factors belong to the σ^{70} family which is generally divided into 4 groups (group 1, 2, 3 and 4) (Lonetto et al., 1992). These groups are defined by structure. Their structure varies by four conserved regions within the four structured domains. Group 1 are primary σ factors transcribing housekeeping genes. Bacteria typically only have one primary σ factor. Primary σ factors have domains $\sigma^{1.1}$, σ^{2}, σ^{3} and σ^{4}. Domain $\sigma^{1.1}$ is unique only to group 1, not allowing the σ factor to bind to DNA without the RNAP core (Hook-Barnard & Hinton, 2009). Some from this group also have a non-conserved region (NCR) (Leibman & Hochschild, 2007). The structure of σ^{70} from *E. coli* can be seen in Fig 2. Groups 2, 3 and 4 are alternative σ factors, primarily involved in transcription of non-housekeeping genes. None of these groups

have $\sigma^{1.1}$. σ factors of group 2 are very similar to group 1, often called primary-like σ factors, but lack $\sigma^{1.1}$ and are non-essential. Group 2 σ factors are involved in stress responses during stationary phase. In *E. coli,* this is σ^{38} (Hengge-Aronis, 1993). Group 3 σ factors typically but not always include σ^2, σ^3 and σ^4 domains. They are involved in responses to general stress and heat stress as well as flagellum biosynthesis and sporulation (reviewed in M. S. Paget & Helmann, 2003). Group 4, also called the extracytoplasmic function (ECF) group, has σ factors with the σ^2 and σ^4 domain. These σ factors are typically involved in responses to signals from outside the cell, like envelope stress or iron transport. Domains σ^2 and σ^4 are present in all σ factors and are highly conserved as they are responsible for binding the RNAP core, the -10 and -35 element and promoter melting (Rhodius et al., 2013). Alternative σ factors are often regulated by anti-σ factors, proteins which typically inhibit the σ factors from binding the RNAP core (reviewed in M. S. Paget, 2015).

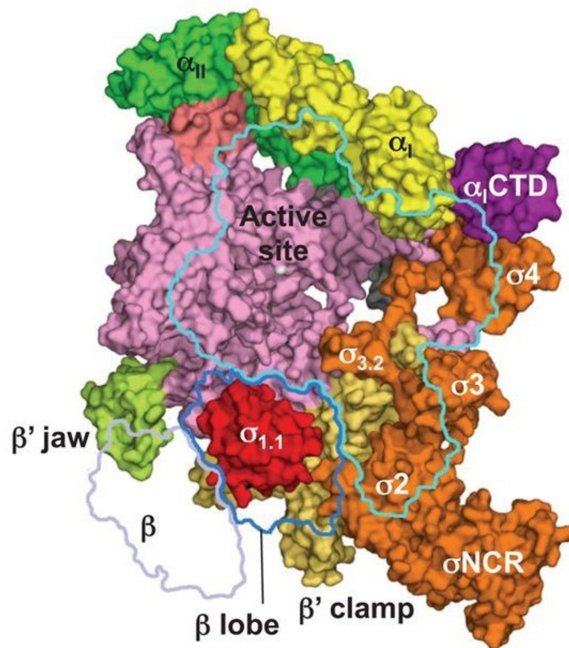

Figure 2. Structure of σ^{70} from *E. coli.* The figure shows a 3D model of the RNAP holoenzyme, focusing on the structure of σ^{70}. The subunits of the RNAP core are labelled. The β subunit is outlined in grey and blue to clearly show the structure of the σ factor. The domains of the σ factor are labelled and in orange and red colour. The structure includes the $\sigma^{1.1}$ domain unique for group 1 σ factors and the non-conserved region (NCR) present in σ^{70}. Adapted and modified from (Murakami, 2013).

3. Bacterial growth curve

As growth of a bacterial culture progresses, changes in the number of bacteria can be observed. The bacterial growth curve was first mathematically analysed by R. E. Buchanan in 1918. He described behaviour of cultures during these so-called life stages in detail. Of all the life stages, today we primarily recognise the lag phase, exponential phase, stationary phase and the death phase. The lag phase occurs early after inoculation into new media and is characteristic by growth slower than the maximum growth rate of the bacteria. The exponential phase follows the lag phase. The bacteria grow at their maximum growth rate and experience the shortest generation time. During this stage, all the bacteria are not limited by resources. When resources

become limiting, the culture begins to enter the stationary phase. The number of bacteria becomes stagnant until the bacteria use-up even more resources and start to die. Dying of the bacteria due to lack of resources is known as the death phase (Buchanan, 1918). The graph with the described growth curve can be seen in Fig 3.

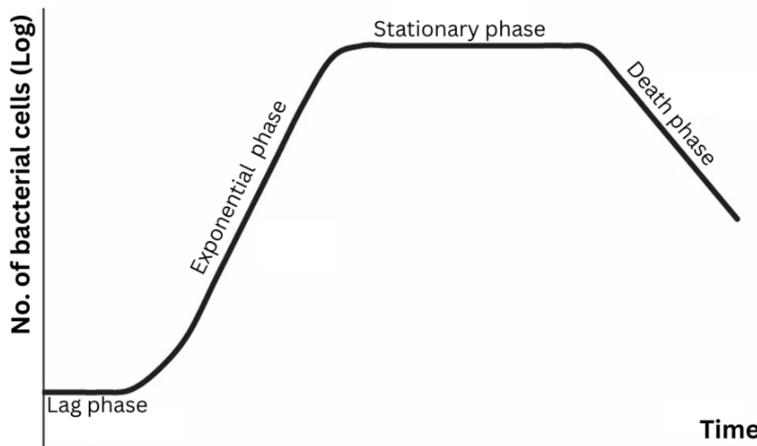

Figure 3. Bacterial growth curve. The figure shows a graph depicting the growth curve of bacteria during cultivation. The x-axis represents the elapsed time, and the y-axis represents the number of bacterial cells in a logarithmic scale. The above-described growth phases are labelled on the graph. Adapted and modified from (Wang et al., 2015).

4. Regulatory proteins binding RNA polymerase

RbpA (RNA polymerase-binding protein A) is a 14 kDa RNAP-binding dimeric protein which has been discovered in *Streptomyces coelicolor* by Paget *et al.,* (2001). The protein has its closest homologues in *Mycobacteria*. RbpA is specific to *Actinobacteria* and therefore regulation with RbpA cannot be found in *E. coli* (M. S. B. Paget et al., 2001). The protein has been deemed a transcription activator which binds the RNAP holoenzyme containing σ^A and σ^B. Regulation of gene expression differs from phylum to phylum and thus it is no surprise that *Proteobacteria* often utilise different mechanisms. For example, *Mycobacterium* species lack DksA which is present in *E. coli* and interfere with synthesis of stable RNA transcripts (Perederina et al., 2004). The DksA protein binds RNA polymerase, while (p)ppGpp (guanosine penta/tetraphosphate) destabilises the open promoter complex during transcription initiation (Paul et al., 2004; Raghavan & Chatterji, 1998).

Functioning as an RNA polymerase-binding transcription factor, CarD is considered an important regulator of rRNA transcription in *M. tuberculosis* (Stallings et al., 2009). Regulation of rRNA transcription plays a major role in tuberculosis persistence since downregulating ribosome synthesis and thus translation is key when resisting the host´s attempts to clear the infection. The bacterium is able to withstand starvation and damage by becoming dormant (Betts et al., 2002). Apart from rRNA regulation, the protein is involved in stabilising the open promoter complex (Davis et al., 2015). While CarD was originally identified in *Myxococcus xanthus*, CarD homologs, CdnL (CarD N-terminal like), have been subsequently identified across many phyla. These identified phyla include *Actinobacteria, Firmicutes, α- and δ- Proteobacteria,*

Chlamydiae, Cyanobacteria, Deinococcus-Thermus and *Spirochaete* (Bae et al., 2015; Nicolás et al., 1994). Regulation by CarD has been characterised also in *Mycobacterium smegmatis*. Furthermore, it is important to highlight that CarD is not found in *E. coli*, hence this regulation has long been overlooked. Research based on the depletion of CarD transcripts shows the expression of CarD in *M. tuberculosis* and *M. smegmatis* is essential for viability as the depletion clearly correlates with cell death (Stallings et al., 2009).

5. Regulatory RNAs binding RNA polymerase

6S RNA is an sRNA (small RNA) which was first discovered in *E. coli (Hindley, 1967)*. As reviewed by Storz *et al.,* (2011), sRNAs are small regulatory RNAs which the scientific community only recently began recognising for the major role they play in regulation of gene expression. Their involvement allows bacteria to adapt to the environment and various stresses since they can vastly affect processes in the cell. Their typical length is 50-300 nt and they can bind mRNA or modulate the activity of proteins. When binding to mRNA, there is some degree of complementarity of the sRNA to the target mRNA. Regulation by sRNAs is post-transcriptional, affecting the stability of mRNA and influencing the efficiency of its translation. When binding proteins sRNAs can resemble other nucleic acids thus changing the activity of the protein (reviewed in Storz et al., 2011)

The secondary structure of 6S RNA resembles the structure of an open promoter (Barrick et al., 2005). 6S RNA thus interacts with the RNAP holoenzyme which has the housekeeping σ factor, σ^{70}. The binding of the sRNA decreases the transcriptional activity of the holoenzyme as it cannot bind to a promoter (Wassarman & Storz, 2000). 6S RNA genes are widely distributed among bacteria. To mention a couple, these RNAs have been identified in many species of the phyla *Firmicutes*, *Cyanobacteria*, *Spirochaetes*, *Bacteroidetes* and *Proteobacteria.* Some phyla even have two 6S RNAs (Wehner et al., 2014).

5.1. Regulatory RNAs binding RNA polymerase specific to *Actinobacteria*

Ms1 is an sRNA (small RNA) present in *M. smegmatis* or *S. coelicolor* as well as in majority of *Actinobacteria* (Vaňková Hausnerová et al., 2022). Despite being recently found when searching for a 6S RNA homolog in *Mycobacteria*, Ms1 does not bind with the RNAP holoenzyme with a housekeeping σ factor (Panek et al., 2011). Instead of binding the RNAP holoenzyme, this sRNA binds to the RNAP core (Hnilicová et al., 2014). Ms1 is about 300 nt long and was first discovered in *M. smegmatis*. MTS2823, an sRNA present in *M. tuberculosis,* is its homolog (Panek et al., 2011).

6. *Actinobacteria* and regulation of gene expression

Actinobacteria is a highly diverse phylum morphologically and ecologically. The phylum includes single-celled species as well as species which form multicellular mycelia. Species of this phylum can be found both in terrestrial and aquatic ecosystems. They show a variety of life strategies as they can be plant symbionts, gastrointestinal commensals as well as plant, animal and human pathogens. Their diversity is also shown on a metabolic level, as they are producers of an enormous amount of bioactive natural products. About two thirds of the antibiotics used are produced by *Actinobacteria*. They show a wide variety of specialised

metabolisms. They are Gram-positive bacteria with genomes of high G+C content, which are often filamentous. Species which form mycelia often sporulate to reproduce. Most of the species of this phylum are aerobic. *Actinobacteria* include the genera *Mycobacterium, Propionibacterium, Nocardia, Corynebacterium, Bifidobacterium, Streptomyces, Frankia* and many more (reviewed in van Bergeijk et al., 2020). Apart from a 16s rRNA gene classification of *Actinobacteria* a genome-based classification has been done. Various phylogenetic trees have been constructed, and the classification of *Actinobacteria* has a substantial history of uncertain relationships within the phylum. Nouioui *et al.,* (2018) proposed that based on their genome-based classification there should be 2 orders, 10 families and 17 genera within the *Actinobacteria* phylum and highlighted the drawbacks of classification solely based on the 16s rRNA gene (Nouioui et al., 2018).

The genus *Mycobacterium* is mainly composed of free-living species which are saprophytic, however the genus includes species which are dangerous pathogens. *M. tuberculosis* and *Mycobacterium bovis* are species which apart from infecting animals have also been identified as obligatory human pathogens. Affecting the skin and peripheral nerves, *Mycobacterium leprae* is another human pathogen from this genus. This species causes leprosy. Many species of this genera are associated with mycobacterial infections. Other examples of species of this genus are *Mycobacterium abscessus* and *Mycobacterium marinum.* The genus *Nocardia* consists of at least 33 pathogenic species mainly associated with opportunistic infections in people who are immunocompromised. These infections lead to nocardiosis which can be widespread and is considered an emerging disease among humans and domestic animals (reviewed in Barka et al., 2016). The genus is also known for producing antibiotics and enzymes. For example, Nargenicin is an antibiotic produced by *Nocardia argentinensis* (Celmer et al., 1980). The genus *Corynebacteria* is important for production of amino acids. As reviewed by Wolf *et al.,* (2017) *Corynebacterium glutamicum* has been intensively genetically engineered for industrial use, now producing over 70 different compounds including biofuels, lactate, succinate and many others. Utilising this species is simple as it is fast-growing, considered safe and produces no endotoxins (reviewed in Wolf et al., 2021). Among other pathogenic species of this genus, *Corynebacterium diphtheria* is a human pathogen which using an exotoxin, the diphtheria toxin, causes acute diphtheria (reviewed in Barka et al., 2016).

As reviewed by Barka *et al.,* (2016) *Corynebacteria, Nocardia* and *Mycobacteria* are mycolic acid-containing genera formerly called the CNM group. Species of the CNM group are characteristic by a waxy cell envelope with mycolic acids. As expected, mycolic acids in the context of *M. tuberculosis* are a long-lived interest to researchers (Minnikin & Polgar, 1966). Viable mutants of *C. glutamicum* that lacked the biosynthesis of mycolic acids were a major step towards studying and understanding the effects of mycolic acids on the growth of such species (Gande et al., 2004). Mycolic acids are also present within the genus *Rhodococcus* (reviewed in Barka et al., 2016).

The genus *Propionibacterium* is associated with human skin. Species of this genus are mainly known for their fermentation as they produce propionic acid. *Proprionibacterium acnes* is considered a commensal but has also been identified as an opportunistic pathogen in relations to acne. The genus *Bifidobacteria* includes species which have probiotic properties and promote health by competitively excluding pathogens.

Bifidobacteria are valuable inhabitants of intestines where they adhere to the intestinal epithelium and produce antimicrobials or bacteriocins (Bevilacqua et al., 2003). As reviewed by Lee & O'Sullivan (2010), due to these benefits, *Bifidobacteria* are often supplemented into food. A decrease in the amount of *Bifidobacteria* correlates with an increase of unwanted *Clostridia* and *E. coli*. The first isolation of *Bifidobacteria* was by Henry Tissier in 1896 from faeces of breast-fed infants. The infants were suffering from diarrhoea when he observed a smaller number of Y-shaped bacteria which he then called "bifid" bacteria (reviewed in Lee & O'Sullivan, 2010). *Frankia* spp. are able to fix atmospheric nitrogen and are symbiotes of actinorhizal plants (Callaham et al., 1978). The species of the genus *Streptomyces* are important producers of antibiotics and other secondary metabolites. They are an important part of soil as they recycle carbon from organic matter. Formation of a mycelium is typical for the genus *Streptomyces,* and it can be observed in the case of *S. coelicolor* which form spores at the tips of their aerial mycelium (Wildermuth et al., 1971). Their mycelia are permanent and well differentiated with a branched morphology (reviewed in Barka et al., 2016).

6.1. Mycobacterium

The genus *Mycobacterium* has been previously mentioned in reference to important pathogens. Thorough research has been and is currently still being done on *M. smegmatis* as unlike the other mentioned *Mycobacteria*, it is a non-pathogenic species. The non-pathogenic nature, fast cultivation and effective genetic manipulation via *E. coli* shuttle plasmids as vectors has been crucial for discovering new information which can be applied when confronting tuberculosis or leprosy (Rawat et al., 2007; Zainuddin et al., 1989).

In *M. smegmatis,* the essential σ factor involved in transcription of housekeeping genes is σ^A (Gomez et al., 1998). Functioning as one of the alternative σ factors, σ^B, is key for transcription during stationary phase and under stress. Nonetheless, studies have discovered σ^B, similarly to σ^A, is also greatly involved in transcription of housekeeping genes during exponential phase (K. Hurst-Hess et al., 2019; Waagmeester et al., 2005). Studies on *M. tuberculosis* by Singha *et al.,* (2023) discovered about 57% of the transcriptome is regulated by σ^A. Of the 57% about 28% were essential genes. According to their RNA-seq analysis, σ^B regulates 433 genes. More than half of these genes were also part of the σ^A transcriptome. However, their study also showed that an overexpression of σ^B is not able to compensate for a missing σ^A (Singha et al., 2023).

6.1.1 RbpA

In *M. tuberculosis*, RbpA binds to the RNAP holoenzyme containing the primary σ factor, σ^A, and the primary-like σ factor, σ^B (Hu et al., 2012a, 2014). The crystal structure of RbpA with the mycobacterial transcription initiation complex by Hubin *et al.,* (2017) can be seen in Fig 4. RbpA consists of 4 domains which are the NTT (N-terminal tail), CD (core domain), BL (basic linker) and the SID (σ-interaction domain). These domains can be also seen in Fig 4. The domains are characteristic by the following: BL binds DNA, SID binds primary and primary-like σ factors, CD interacts with β′ and the NTT interacts with β′and σ (Hubin et al., 2017). Research by Prusa *et al.,* (2018) showed that while the SID domain is enough for interacting with the RNAP, having dysfunctional domains reflects in the growth of *Mycobacteria*. In *M.*

smegmatis having functional domains was essential for optimal growth and in *M. tuberculosis* for viable growth. I*n vitro* experiments showed the BL and SID domains were necessary for the stabilisation of the open promoter complex (RPo). The NTT and CD domains opposed the stabilisation. The stabilisation of the RPo relied heavily on interactions of RbpA with promoter DNA and the σ factor. As it could be expected, results showed the presence of dysfunctional BL and SID domains, which stabilise RPo had a more severe effect on growth than the NTT and CD domains (Prusa et al., 2018).

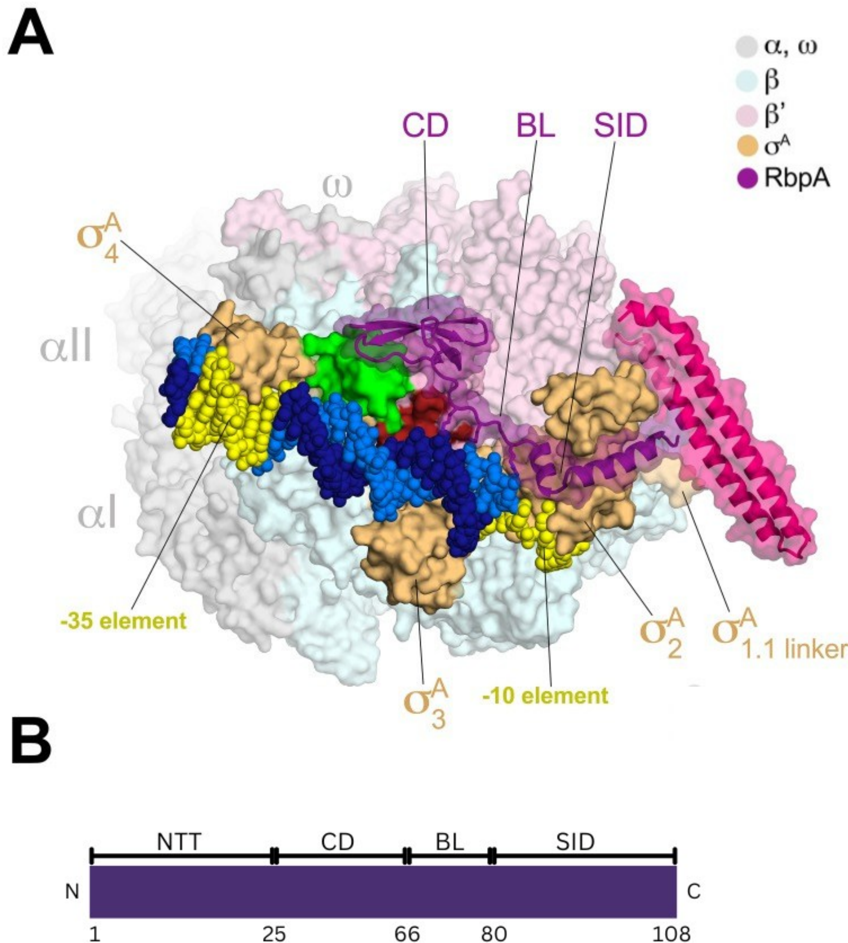

Figure 4. Structure of mycobacterial transcription initiation complex with *M. smegmatis* RbpA and a representation of the domain structure of *M. smegmatis* RbpA. The structure of the mycobacterial transcription initiation complex with RbpA from *M. smegmatis* is seen in (A). Parts of the RNAP holoenzyme are colour-coded according to the legend while RbpA is in dark purple. Some of the mentioned RbpA domains are labelled in (A). The representation of the domain structure of *M. smegmatis* RbpA is seen in (B) and it is labelled with all the mentioned secondary structure elements: N-terminal tail (NTT), core domain (CD), basic linker (BL) and the σ-interaction domain (SID) Adapted and modified from (Hubin et al., 2017).

As researched by Hu *et al.,* (2012) in *M. tuberculosis* the interaction of RbpA with the RNAP holoenzyme is the most prominent during a stress response and during stationary phase. The presence of RbpA regulates the formation of the σ^A-containing RNAP holoenzyme. When RbpA binds, a change of the RNAP core structure occurs and the interaction of σ^A and the RNAP core becomes more stable, increasing

transcription activity. The amount of transcriptionally active promoter complexes increases and the affinity of σ^A to the RNAP core is higher in the presence of RbpA. In the absence of RbpA, the active promoter complexes are unstable and less transcriptionally active compared to when RbpA is present. Binding of RbpA leads to more transcriptionally active promoter complexes. The binding site of RbpA is located on the β subunit, outside of the RNAP main channel. The other proposed function of RbpA is optimising the interaction of the σ factor with the RNAP core when forming the holoenzyme (Hu et al., 2012b). As transcription initiation occurs, RbpA does not dissociate from the promoter complex and continues to be a part of the initiation complex even after an RPo has been formed (Hu et al., 2014).

While the function of RbpA was first observed involving σ^A, further exploration by Hu *et al.,* (2014) uncovered a significant interaction between RbpA and σ^B. They found that in *M. tuberculosis*, the RNAP holoenzyme containing σ^B severely lacks the ability to melt housekeeping promoters and to form stable open complexes even when a perfect -10 consensus element was present. For an open complex to form and thus efficient transcription, RbpA was required. The reliance of σ^B on RbpA was greater than that of σ^A on RbpA.Thus it was concluded that transcription of housekeeping genes by σ^B-RNAP is much more heavily regulated by RbpA than transcription by the primary σ^A (Hu et al., 2014). As previously mentioned in chapter 3.1, despite its major role during stationary phase and under stress, σ^B has been identified to take part in transcribing housekeeping genes (K. Hurst-Hess et al., 2019; Waagmeester et al., 2005). When considered altogether, regulation by RbpA emphasises the role σ^B plays in times of stress. Hu *et al.,* (2014) further proposed that this could be due to σ^B possibly overtaking expression of housekeeping genes when σ^A is unavailable, RbpA-dependent regulation could have a huge impact on regulation of gene expression under stress. In contrast, other alternative σ factors which are not primary-like (group 3 and group 4) are capable of promoter melting without the presence of RbpA (Hu et al., 2014).

Since the presence of RbpA is associated with rifampicin resistance in *M. smegmatis* and in *S. coelicolor*, its effect on rifampicin was studied (Dey et al., 2010; Newell et al., 2006). Rifampicin binds in a pocket of the β subunit of the RNAP core. As the antibiotic binds it is located deep within the DNA/RNA channel, but far away from the RNAP active site. Based on the nature in which rifampicin binds, the antibiotic likely inhibits RNAP by obstructing the path of the elongating transcript after reaching a length of 2 to 3 nucleotides during transcription. Binding of the antibiotic does not allow the RNAP to synthesise phosphodiester bonds as the path is obstructed (Campbell et al., 2001). Focusing on how RbpA affects rifampicin sensitivity in *M. tuberculosis*, Hu *et al.,* (2012) presented that in the presence of RbpA, the amount of yielded transcripts was larger even when transcription was still being greatly inhibited by rifampicin. This finding showed the effect RbpA has on rifampicin resistance is indirect as RbpA does not affect the binding of rifampicin (Hu et al., 2012b). With the persisting issue of bacterial resistance to antibiotics understanding the behaviour and effects of RbpA in *M. tuberculosis* is of great importance.

Apart from rifampicin, the effect on fidaxomicin sensitivity in *M. smegmatis*, was studied by Prusa *et al.,* (2022). Fidaxomicin inhibits RNAP transcriptional activity by binding to RNAP and keeping the RNAP clamp from closing, not allowing the RNAP to secure promoter DNA in the active site. The closing of the RNAP clamp happens when the RPo is forming (Boyaci et al., 2018). As mentioned previously, the

stabilisation of RPo by RbpA relies heavily on RbpA interacting with promoter DNA and the σ factor (Prusa et al., 2018). Extending this finding, decreased affinity of RbpA to promoter DNA and the σ factor led to increased sensitivity to fidaxomicin. The NTT domain partakes significantly in sensitivity to fidaxomicin. Sensitivity to fidaxomicin is not only relevant to RbpA, but also to CarD, which similarly to RbpA stabilises RPo (Prusa et al., 2022).

As mentioned previously, the evolution of RNAP in *Actinobacteria* is divergent to RNAP in *Proteobacteria* and thus unique (Lane & Darst, 2010). Based on analysis of the constructed phylogenetic trees, Dey *et al.,* (2012) observed a co-evolutionary link between the evolution of actinobacterial RNAP (specifically the β and β′subunits) and RbpA. This observation is consistent with RbpA homologues being found only in *Actinobacteria* as stated in chapter 4 (M. S. B. Paget et al., 2001). The co-evolutionary link highlights the importance of RbpA during the evolution of actinobacterial RNAP (Dey et al., 2012).

6.1.2 CarD

An RNA polymerase-binding transcription factor which can be found in *M. tuberculosis* or *M. smegmatis*, CarD, stabilises RPo (Davis et al., 2015). It also takes part in a global regulatory response by regulating rRNA transcription (Stallings et al., 2009). Based on the structures by Gulten & Sacchettini (2013), the CarD protein in *M. tuberculosis* consists of two domains, the N-terminal domain and the C-terminal domain, which are connected by an α-helix and loop. The N-terminal domain is labelled CarD-RID (CarD-RNAP interacting domain) and interacts with the β1 domain of the β subunit. (Gulten & Sacchettini, 2013). The CarD-RID is similar to the RID of the transcription repair coupling factor (TRCF) found in *E. coli* (Deaconescu et al., 2006). The C-terminal domain is labelled CarD-CTD (CarD-C-terminal domain) and consists of α-helixes (Gulten & Sacchettini, 2013). The CarD-CTD includes a conserved tryptophan residue (W86) which is important for the function of CarD (Srivastava et al., 2013). The structure can be seen in Fig 5. Based on structure models by Bae *et al.,* (2015) it was concluded that *Thermus aquaticus* CarD interacts with the upstream ds/ssDNA junction of the transcription bubble via the CarD-CTD. The CTD is positioned in such a manner because of the binding of CarD-RID to the RNAP holoenzyme. The side chain of the conserved tryptophan residue (W86) of the CarD-CTD is wedged into the minor groove of the upstream edge of the ds/ssDNA junction. Strains with mutated W86 showed compromised CarD activity in comparison to wild-type strains (Bae et al., 2015). CarD as a whole has been observed to be highly conserved in other mycobacterial species such as *M. leprae* (Stallings et al., 2009).

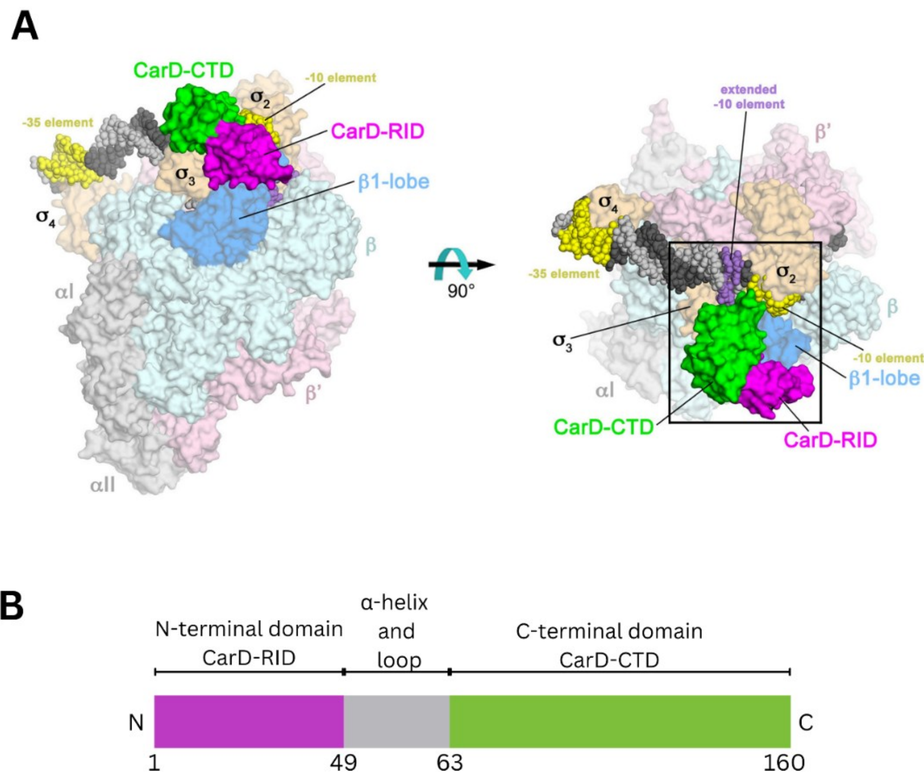

Figure 5. Crystal structures of *T. aquaticus* CarD/RPo complexes and a representation of the domain structure of *M. tuberculosis* CarD. (A) shows the structure of CarD in complex with a transcription initiation complex (RPo). Both the N-and C-terminal domains are labelled and shown in bright green and magenta colours. The RPo is also labelled with parts of the RNAP holoenzyme (α, β and σ subunits) and the promoter DNA (-35, -10 and extended -10 element). The image also depicts how CarD interacts with the β1 domain and the ds/ssDNA junction of the transcription bubble. The β1 domain corresponds to residues 18-138 and 333-392 of the β subunit. The image offers two different views of the complexes. Adapted and modified from (Bae et al., 2015). (B) depicts a representation of the structure domains based on data by Gulten and Sacchetinni (2013). The N- and C- terminal domains are colour-coded according to (A) in green and magenta. (B) includes the connecting part between the two terminal domains made by the α-helix and the loop. The connecting part is in grey colour. The previously mentioned secondary structure elements are labelled (Gulten & Sacchettini, 2013).

According to early studies by Stallings *et al.,* (2009) expression of CarD is associated with DNA damage and starvation and is necessary for viability of *M. tuberculosis* and *M. smegmatis*. Using genotoxins which damage dsDNA and starvation in PBS (phosphate buffered saline) the researchers observed CarD mRNA is upregulated in response to genotoxic stress and nutrient deprivation. During oxidative stress and starvation, CarD depleted cells accumulated rRNA many-fold more in comparison to what was observed in wild-type cells. The protein is not only associated with stress responses, but its downregulation is linked to slow strain growth and is essential. Using whole-genome profiling they identified that CarD depletion correlated with an upregulation of rRNA genes (Stallings et al., 2009).

Despite the above-described data, CarD is now regarded as an rRNA transcription activator rather than an rRNA transcription inhibitor. Regarding CarD as an activator arose with new data which showed that the protein stabilises RPo on rRNA genes. Research done by Davis *et al.,* (2015) analysed RPo formation from a *M. tuberculosis* rRNA promoter. They compared RPo formation with RNAP from *E. coli* and RNAP from *M. bovis*. While RNAP from *E. coli* formed stable and irreversible promoter complexes, mycobacterial RNAP activity in the absence of CarD led to very unstable and reversible promoter complexes. Furthermore, while *E. coli* RNAP transcription activity was high, mycobacterial RNAP activity was comparable only after the addition of CarD. CarD was identified to stabilise mycobacterial RNAP promoter complexes. Since the researchers hypothesised that CarD stabilises RPo by not allowing the transcription to collapse by reannealing together, they conducted further experiments. Using a synthetic promoter template which would form a transcription bubble unable to collapse, they found the activity of the mycobacterial RNAP was comparable to that of *E. coli* RNAP and adding CarD has no distinct effect on the half-life of the RPo (Davis et al., 2015). To further understand how the mechanism by which CarD functions, Rammohan *et al.,* (2015) used *M. bovis* and showed that CarD functions in a concentration-dependent manner and with different affinities binds open or closed holoenzyme promoter complexes (Rammohan et al., 2015).

With clarifying CarD regulation under starvation in mind, Li *et al.,* (2022) used the same *M. smegmatis* strain mc^{2}155 as was used by Stallings *et al.,* (2009). They found that although starvation correlated with an upregulation of CarD mRNA, it also correlated with a decrease in protein levels of CarD. As seen in Fig 6, like in the starvation experiment, a decrease in CarD levels was observed during stationary phase. Their further experiments showed that these correlations can be observed also in *M. tuberculosis* and *M. bovis* strains, thus suggesting it is common in *Mycobacteria*. Using host-like stresses, low pH and a hypoxic environment, they observed the same behaviour of CarD protein levels as during starvation and stationary phase. All the presented data would indicate CarD protein levels decrease as a reaction to stress in efforts to slow down translation and metabolic rate. The decrease is regulated by a mechanism which uses antisense RNA and Clp protease. The Clp protease degrades the CarD protein during stationary phase and starvation. The antisense RNA named AscarD, is located on the antisense strand of the operon from which CarD is transcribed. AscarD is transcribed by the alternative σ factor, σ^F, in response to starvation. The expression of AscarD was shown to inhibit the synthesis of CarD (Li et al., 2022a). Regulation by CarD is a complicated dispute of data which tends to point in opposite directions. Research by Zhu *et al.,* (2022) gives an insight as to why it is so problematic to define the role of CarD. Their research showed mycobacterial CarD can upregulate transcription from some promoters while downregulating it from others. To conclude, CarD can take the role of both an activator and inhibitor (Zhu et al., 2019).

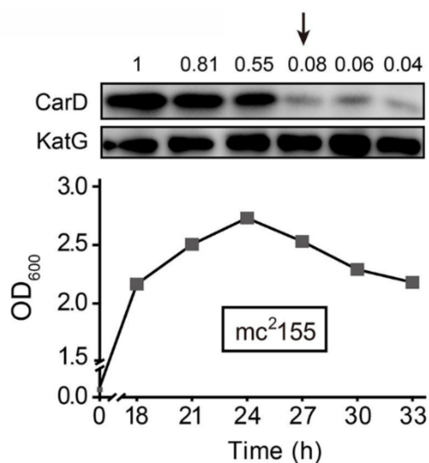

Figure 6. Growth curve of *M. smegmatis* strain mc^{2}155 and western blot results of CarD protein levels. On the bottom of the image, we can see the growth curve of the mycobacterial strain with the respective OD measured at wavelength of 600 nm. After 24 hours, we see a decrease of growth, following a typical exponential curve, thus signalling a transition from an exponential growth phase to a stationary growth phase. The square symbols on the curve mark the sampling times during the growth of the strain. On top of the graph, we can see the western blot results of the CarD protein levels of growth of the same strain. The intensity of each band was normalised to the loading control (KatG). Setting the normalised value of the first sample to 1, the following values represent fold changes relative to the first sample. The arrow above the fourth band with a relative value of 0.08 marks a sharp decrease in CarD protein levels. The decreasing gradually continues as stationary phase progresses. As the loading control they used the protein KatG, a catalase-peroxidase. Adapted from (Li et al., 2022).

Since research about CarD proved the presence of CarD can have immense effects on gene expression, it comes as no surprise that further research targeted the interaction between CarD and the RNAP β-subunit. Weiss *et al.,* (2012) identified point mutations which lead to the weakening of this interaction. The amino acid sequence of the β1 domain from RNAP and the CarD-RID was mutated, substituting only chosen residues. On the β1 domain they conducted research using two substitutions. One of the substitutions was E138 to R138, substituting glutamic acid with arginine at position 138. The other substitution was I147 to A147. On the CarD-RID they conducted research also using two substitutions. The substitutions were R25 to E25 and R47 to E47. The mutations are associated with a slower growth rate and have been showed to compromise the survival of the mutated strains. Studied *M. tuberculosis*, strains with these described point mutations were more sensitive to rifampicin as their MIC (minimum inhibition concentration) decreased significantly. In some cases, even strains previously resistant to rifampicin were more sensitive to oxidative stress when the interaction was weakened by these mutations. When examined further, simply depleting CarD did not lead to the increased sensitivity to rifampicin; therefore, it is the interaction of CarD with the β subunit which should be targeted by rifampicin rather than the protein alone. In *M. smegmatis* the effects under different antibiotics varied. Unlike observed with rifampicin, when streptomycin was used both depleted CarD and the weakened interaction led to increased sensitivity to the antibiotic (Weiss et al., 2012).

Taking a different approach, Garner *et al.,* (2017) focused on how increasing the affinity of CarD for RNAP affected the growth, rRNA transcription and virulence of *M. tuberculosis*. Based on crystal structures of the CarD/RNAP complex, they identified possible substitutions in the amino acid sequence which would increase affinity of CarD for RNAP. They constructed strains mutated in the CarD-RID. In $CarD^{I27F}$ and $CarD^{I27W}$ mutants, they observed stabilising of the RPo at lower concentration of CarD than in wild-type stains. The increased affinity also led to an increase in initial RNA product formation. Despite mutant strains having the same amount of CarD mRNA as the wild-type strains, their amount of CarD protein was larger (Garner et al., 2017). This was explained by the CarD degradation mechanism by Clp protease mentioned above, since CarD bound to RNAP would be protected from this degradation (Li et al., 2022). Surprisingly, an increase in growth rate, but not a change in rRNA transcription was observed, thus presenting a case where the two are uncoupled in *M. tuberculosis*. Increased affinity resulted in a loss of virulence and thus the level of affinity is important for the bacteria to optimise their virulence (Garner et al., 2017).

Further expanding on the role of CarD in *Mycobacteria*, the protein has interacting partners with affect gene expression in their own characteristic ways. One of the partners, found by Shoman *et al.,* (2024) in *M. smegmatis,* is CrsL (CarD RNA polymerase small linker), a recently discovered small protein bound directly to CarD both in exponential and in stationary phase. The interaction of the two proteins, forming a CarD-CrsL complex, appears to be mainly regulated by the levels of CarD and CrsL which change depending on the stage of cell growth. CrsL itself seems to be regulated post-transcriptionally as crsL mRNA depletion using strains with a CRISPR system does not reflect in the amount of CrsL protein present. Like CarD, CrsL tends to interact mainly with actively transcribed genes and the proteins likely cooperate to regulate gene expression. Apart from this regulation, CrsL allows for growth at higher temperatures. Many homologues of CrsL have been identified including in *M. tuberculosis* (Shoman et al., 2024).

6.1.3 HelD

A helicase-like factor, HelD, first discovered in *Bacillus subtilis* and later in *M. smegmatis*, similarly to other proteins involved in regulating gene expression interacts with RNAP (Carrasco et al., 2001; Delumeau et al., 2011). HelD-like sequences were searched for by Larsen et al., (2021) and it was concluded HelD is mainly among the phyla *Firmicutes* and *Actinobacteria* which are both Gram positive eubacteria. This phylogenetic research allowed for a third class of HelD to be identified in *Deltaproteobacteria* which are Gram negative (Larsen et al., 2021). Prior to this study, there were only two identified classes of HelD which show similar function but differ in their structure and mechanism. The first identified HelD from *Bacillus subtilis* is class I, while HelD from *M. smegmatis* is class II (Kouba et al., 2020; Newing et al., 2020). The newly constructed phylogenetic tree of HelD showed class I is mainly present in *Firmicutes*, *Clostridia* and *Bacteroidia* and class II in *Actinobacteria* and *Acidimicrobilia*. Multiple HelD proteins can be present at once (Larsen et al., 2021). Despite being a transcription factor, unlike the others previously discussed, HelD is primarily involved in the recycling of RNA polymerase (Wiedermannová et al., 2014). The protein binds to the RNAP core, although a transitional complex of the RNAP holoenzyme with HelD has been observed as well. HelD is essential in freeing stalled RNAP by helping clear the RNAP of any nucleic acids. The protein can also

free RNAP following an unsuccessful dissociation of the RNAP after transcription termination (Kouba et al., 2020).

According to cryo-EM structures by Kouba *et al.,* (2020), the protein has a so-called crescent-like shape. The structure and binding of the protein is essential for its function, as one of the ends extends into the primary channel while the other extends into the secondary channel of the RNAP core. The N-terminal domain of the protein initiates the binding by interacting with the secondary RNAP channel. The Pch-loop (primary channel loop) and the CO-domain (clamp-opening domain) extend into the primary RNAP channel, leading to the dilation of the RNA exit channel. By this mechanism any nucleic acids leave from the active site of the RNAP. To further expand on the mechanism, the Pch-loop folds within the primary channel and binds to the active site of the RNAP core. The CO-domain secures the primary channel in its wide-open state. It is the interaction with HelD that leads to the previously closed primary channel opening. The Pch-loop and the CO-domain are part of the HelD-specific insertion domain as seen in Fig 7. The structure of the RNAP core and how HelD from *M. smegmatis* binds to it can also be seen in Fig 7. (Kouba et al., 2020).

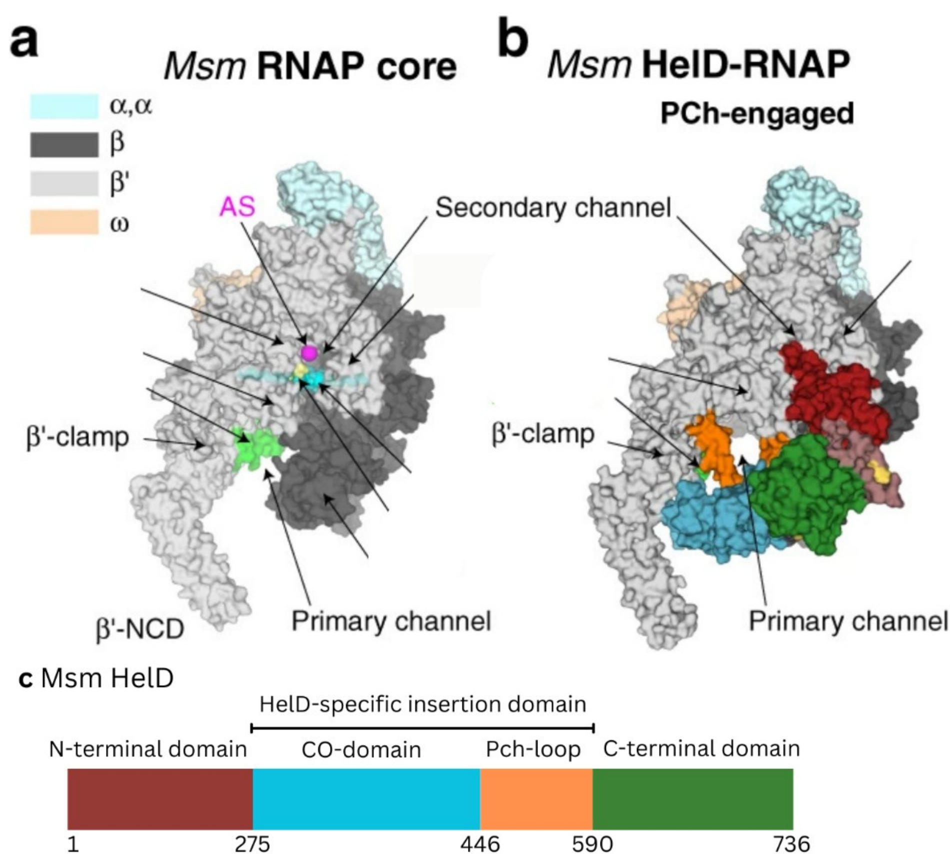

Figure 7. Structures of *M. smegmatis* RNAP core and *M. smegmatis* HelD-RNAP complex by cryo-EM and a representation of the domain structure of *M. smegmatis* HelD. The RNAP core alone is seen in (a), the Msm HelD-RNAP complex is seen in (b). The subunits of the RNAP core (a) have an associated legend while the domains of HelD in (b) are colour-coded according to the depicted domain structure of HelD in (c). The representation of the domain structure of Msm HelD is labelled with secondary structure elements: N-terminal domain, CO-domain and Pch-loop and C-terminal domain. As previously mentioned, the HelD-specific insertion domain consists of the CO-domain and the Pch-loop. Adapted and modified from (Kouba et al., 2020).

Recent research by Kovaľ *et al.,* (2024) proved mycobacterial HelD forms complexes with the RNAP holoenzyme containing σ^A and with the previously mentioned RbpA. CarD is not present in these complexes as the binding of HelD leads to conformational changes of the complex, not allowing CarD to bind. The way HelD is released after binding to RNAP was long unknown. Recent studies identified the mechanism by which HelD is released and RNAP can thus again engage in transcription initiation. Binding of ATP and partially GTP for hydrolysis occurs at the HelD NTPase unit. The binding and hydrolysis promote the release of HelD from RNAP. HelD was also found to protect *M. smegmatis* from rifampicin treatment by binding to RNAP and changing the typical binding site of rifampicin (Kovaľ et al., 2024).

Sensitivity to rifampicin is affected also in other species. As investigated by Hurt-Hess *et al.,* (2022) in *M. abscessus,* the HelD homolog, *MAB_3189c,* takes part in resistance to rifampicin. The deletion of the gene led to increased sensitivity to rifampicin and rifabutin, a derivate of rifampicin which has proven effective against *M. abscessus* (Aziz et al., 2017). An intrinsic resistance to rifampicin is typical for this species due to an ADP-ribosyltransferase (Rominski et al., 2017). *MAB_3189c* was thus renamed by the researchers to *Mab_helR*. A rifampicin-associated element (RAE) that is typically associated with enzymes inactivating rifampicin, was identified upstream of the gene and is responsible for regulating the expression of *Mab_helR* (Spanogiannopoulos et al., 2014). The PCh-loop of this HelD homologue is involved in the rifampicin resistance (Hurst-Hess et al., 2022). As HelD is widely distributed among *Actinobacteria*, a HelD homologue, HelR, has been identified by Surette *et al.,* (2022) in *Streptomyces venezuelae*. HelR was identified thanks to the presence of a RAE. Similarly to the previous study, they found HelR also confers resistance to rifamycin antibiotics such as rifampicin, rifabutin or rifaximin. The expression of HelR was found to be rifamycin inducible (Surette et al., 2022).

6.1.4 MoaB2

Regulating gene expression of housekeeping genes can also occur through an interaction of σ^A, the primary σ factor, with a newly identified transcription factor MoaB2. Researched by Brezovská *et al*. (2024) in *M. smegmatis*, the σ^A-MoaB2 interaction inhibits σ^A-dependent transcription. The protein has 17.9 kDa and has two *Escherichia coli* homologues; MoaB and MogA (Kozmin & Schaaper, 2013; Nichols & Rajagopalan, 2002). MoaB is also present in *Pyrococcus furiosus* (Bevers et al., 2008). Using the structure of *M. marinum* MoaB2 and purified *M. smegmatis* MoaB2 crystals, the structure of MoaB2 was determined. The researchers revealed a hexamer with 6 subunits each consisting of a central β-sheet surrounded by seven helices. The hexamer is made of two trimers as seen in Fig 8. The protein is also a hexamer in solution, like other identified MoaB2 proteins (Brezovská et al., 2024).

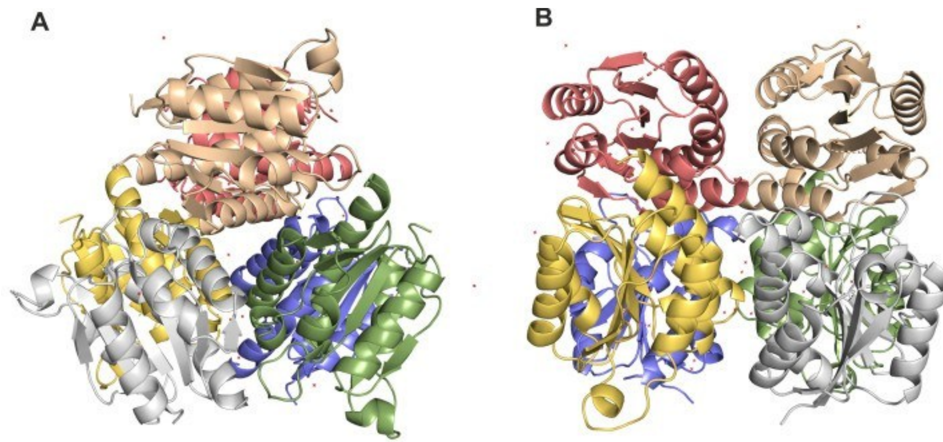

Figure 8. 3D structure of MoaB2 from *M. smegmatis*. (a) shows the hexamer with the 6 subunits each formed by the central β-sheet surrounded by seven helices. Each subunit has a different colour. (b) shows the same hexamer but from a different angle, allowing a better view of the two trimers which form the MoaB2 hexamer. The subunits in (b) match their colour in (a). (a) is a top view of MoaB2. (b) is a side view of MoaB2. Adapted from (Brezovská et al., 2024).

Sequestering σ^A by MoaB2 results in σ^A not being able to interact with the RNAP core and thus form a holoenzyme and initiate transcription of housekeeping genes. Using immunoprecipitation pull-down experiments, it has been identified, that MoaB2 is present during both exponential and stationary phase. Furthermore, MoaB2 was found to not bind the RNAP holoenzyme, meaning it does not interact with σ^A when it is in complex with the RNAP core. MoaB2 does not interact with the tested alternative factors, σ^B, σ^E, σ^F, σ^H and σ^G. This lack of interaction supports the findings where MoaB2 was found to inhibit σ^A-dependent transcription but not σ^B-dependent transcription. Formation of the MoaB2-σ^A complex can thus modulate σ^A dependent transcription *in vitro.* There is clear reasoning as to why this is the case since alternative σ factors, including σ^B, do not contain the unstructured N-terminal domain of σ^A, $\sigma^A{}_N$. Building on this idea and using NMR spectrometry, the interaction between MoaB2 and σ^A has been thoroughly analysed. It has been identified that the C-terminal part of the $\sigma^A{}_N$ domain is responsible for the interaction which occurs during their binding. Size exclusion chromatography experiments showed MoaB2 is *in vitro* capable of binding to σ^A without the need of any additional factors. Apart from inhibiting transcription initiation, its ability to stabilise σ^A in the cell has been researched *in vivo* using MoaB2 depleted strains prepared by CRISPR Cas9. MoaB2 depleted strains were also used during growth experiments which showed MoaB2 in not essential *in vivo* and the lack of MoaB2 does not affect growth of these modified strains (Brezovská et al., 2024).

6.1.5 Ms1 RNA

Whereas MoaB2 interacts with a σ factor, Ms1, an sRNA, uses a different mechanism to regulate gene expression. As identified by Hnilicová *et al.,* (2014), Ms1 in *M. smegmatis* binds to the RNAP core. Interestingly, they did not identify any highly expressed sRNAs which would bind to the RNAP holoenzyme with σ^A. Ms1 is expressed in both exponential and stationary phase but mainly in stationary. Bioinformatic analysis predicted an open promoter-like secondary structure of Ms1. To give more detail, Ms1 consists of a

long hairpin which has a single stranded central bubble as seen in Fig 9. The central bubble was shown to be important for the interaction of Ms1 with the RNAP core as deleting the central bubble led to a significant decrease in binding (Hnilicová et al., 2014). It appears Ms1 only interacting with RNAP core connects to findings where RNAP core without a σ factor is able to bind to DNA templates with a bubble since the presence of the bubble substitutes for otherwise necessary promoter melting (Fredrick & Helmann, 1997).

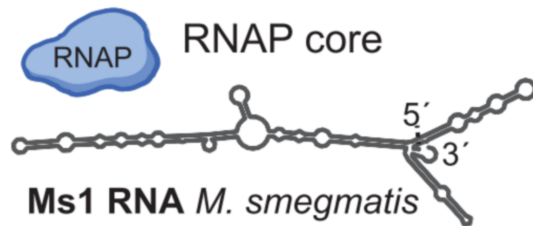

Figure 9. Scheme of the interaction between Ms1 from *M. smegmatis* and the RNAP core. There is a representation of the open promoter-like secondary structure of Ms1 sRNA. The 5´ and the 3´ends are labelled in the figure. The single stranded central bubble of the long hairpin is also depicted**.** In blue, there is the representation of RNAP core without a σ factor as Ms1 does not bind the RNAP holoenzyme. Both the enzyme and the sRNA are labelled. Adapted from (Hausnerová et al., 2024).

Further characterising Ms1 sRNA, Šiková *et al.,* (2019) identified the Ms1 promoter and effects of Ms1 on growth. The identified promoter sequence of Ms1 in *M. smegmatis* is conserved in the promoter sequence of the Ms1 homolog, MTS2823, in *M. tuberculosis*. They also share the same transcription start site. The *M. smegmatis* Ms1 sRNA is transcribed from only one promoter. The sRNA increases about 115-fold in stationary phase, but the activity of the promoter increases only about 2-fold in *M. smegmatis*. To analyse this discrepancy, conducting further experiments showed it is due to the accumulation of Ms1 during stationary phase. As seen in Fig 10, the accumulation occurs because of increased stability of Ms1 in stationary phase (half-life of about 8h) compared to its low stability during exponential phase (half-life of 8 min). Ms1 levels also influence the amount of RNAP present. In Ms1 depleted strains it was observed that the levels of the β and β´ subunits decreased by 2-fold during stationary phase, meaning the presence of Ms1 leads to higher levels of RNAP. This affects the bacteria during outgrowth in fresh media following stationary phase. The increased amount of RNAP due to the presence of Ms1 is likely the reason for the accelerated outgrowth compared to Ms1 depleted strains. Besides the increase of Ms1 during stationary phase, an increase of transcription of the Ms1 sRNA was identified during starvation (growth in PBS). Effects of various stresses were tested. Like other treatments, ethanol treated cells showed an increase in transcription of Ms1. Based on these findings it was concluded that an increase of Ms1 is not a response to general stressing of the cells. Ms1 is not essential as strains lacking this sRNA are viable (Šiková et al., 2019).

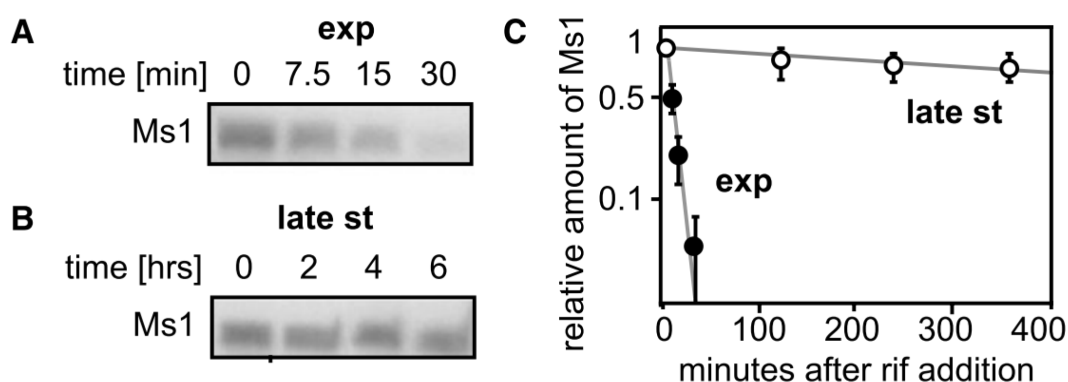

Figure 10. ***M. smegmatis*** **Ms1 stability northern blot results and graph.** (a) and (b) show the northern blot results of the experiment testing the stability of Ms1 in cells during exponential and stationary phase. To observe the half-life of the Ms1 sRNA they treated the cells with rifampicin to stop *de novo* transcription and then isolated total RNA at specific time points. The time points are labelled above the bands and are identical to the time passed since the rifampicin treatment. For (a) they labelled the time passed in minutes due to the short half-life of Ms1 during exponential phase. For (b) they labelled the time passed in hours due to the long half-life of Ms1 during stationary phase. By assessing the intensity of the bands, we can see that during exponential phase Ms1 was quickly degraded (half-life 8 min) while during stationary phase Ms1 was very stable (half-life about 8 hrs). (c) shows a graph of averaged data. The x-axis is the time passed since the rifampicin treatment in minutes, while the y-axis is the relative amount of Ms1. The filled-in circles on the graph represent the progression of the Ms1 degradation during exponential phase. The blank circles on the graph represent the progression of the Ms1 degradation during late stationary phase. According to the total Ms1 sRNA degradation over time indicated by the graph, Ms1 is much more stable during stationary phase than during exponential phase. Adapted from (Šiková et al., 2019).

6.2. Streptomyces

The important antibiotic-producer, *S. coelicolor*, is a species of this genus and like other bacteria uses σ factors to regulate gene expression. A total of 65 genes were identified to encode σ factors (Bentley et al., 2002). The primary housekeeping σ factor is HrdB encoded by the *hrdB* gene (Tetsuo et al., 1991). Apart from HrdB this species has primary-like σ factors which are nonessential: HrdA, HrdC and HrdD (Buttner et al., 1990; Buttner & Lewis, 1992). Another nine alternative σ factors which are σ^B-like and are involved in osmotic stress responses and morphological differentiation including σ^B, σ^F, σ^G, σ^H, σ^I, σ^K, σ^L, σ^M and σ^N. This species also has 51 ECF σ factors (reviewed in Rebets et al., 2018). For example, σ^R is responsible for the cell's response to oxidative stress (M. S. B. Paget et al., 1998).

6.2.1 Ms1 RNA

The Ms1 sRNA discovered in *M. smegmatis* also regulates gene expression in *S. coelicolor*. At first, the *Streptomyces* Ms1 homolog was described by Bobek *et al.,* (2021) as the 6S-like scr3559 RNA. They

identified that scr3559 accelerates the development and increases antibiotic production of *S. coelicolor*. Additionally, they observed the sRNA levels were increasing when transitioning into stationary phase (Bobek et al., 2021). The scr3559 has a 6S-like secondary structure and is 192 nt long, thus when first discovered, scr3559 was even labelled as 6S RNA in this species (Mikulík et al., 2014). The secondary structure can be seen in Fig 11. Recent research further clarified that scr3559 sRNA is indeed a Ms1 sRNA homolog (Hausnerová et al., 2024).

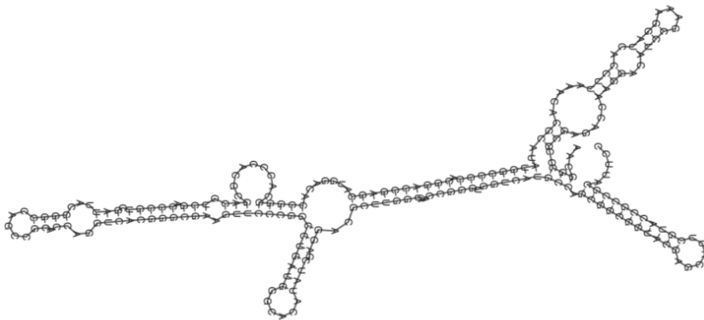

Figure 11. Predicted secondary structure of *S. coelicolor* scr3559/Ms1. The image shows the predicted secondary structure of the Ms1 homolog. In the middle of the 6S-like structure is the central bubble similar to what was predicted in Ms1 sRNA in *M. tuberculosis*. The image can be compared to the representation of Ms1 shown previously. Adapted and modified from (Vaňková Hausnerová et al., 2022).

At first, Vaňková Hausnerová *et al.,* (2022) focused on identifying an Ms1 sRNA homolog in *S. coelicolor*. They then used the same approach and identified other Ms1 homologs in 824 species of *Actinobacteria*. Using *in vivo* experiments, they showed that the identified *S. coelicolor* scr3559/Ms1 sRNA does not bind the RNAP holoenzyme containing HrdB, similarly to what was observed with 6S sRNA in *E. coli*. Instead, they observed it binds only the RNAP core, similarly to what was characterised with Ms1 in *M. smegmatis* (Vaňková Hausnerová et al., 2022). Their following research confirmed scr3559 in *S. coelicolor* is a Ms1 homolog and allowed them to discover a new sRNA. Identifying an 85 nt long sRNA which is particularly short and binds to RNAP during stationary phase. They named it scr0792 according to its transcription locus (Hausnerová et al., 2024).

6.2.2 RbpA

As mentioned previously, the RbpA protein was first discovered in *S. coelicolor* being coeluted with RNAP during gel filtration (M. S. B. Paget et al., 2001). The role of RbpA in gene expression in *M. smegmatis* and *M. tuberculosis* was described previously in chapter 6.1.1. Research by Newell *et al.,* (2006) characterised RbpA in *S. coelicolor*. They began by confirming RbpA binds to RNAP. Constructing RbpA deficient strains, they observed slow growth compared to wild-type strains. Further *in vitro* experiments using an rRNA promoter suggested the absence of RbpA could decrease ribosome biogenesis causing the bacteria to grow slower. The mutant strains also showed increased sensitivity to rifampicin. According to the observed RbpA mRNA levels, rifampicin induces transcription of RbpA despite being a RNAP inhibitor (Newell et al., 2006). In this species, the RbpA gene is part of the σ^R regulon (M. S. B. Paget et al., 2001). In mutant

strains lacking σ^R, they observed RbpA was still expressed and thus expression of RbpA must be partially under control of a different σ factor. Rifampicin was found to induce not only the σ^R-dependent RbpA promoter, but also other promoters transcribed by σ^R (Newell et al., 2006).

6.3. Corynebacterium

As mentioned previously, the *Corynebacteria* genus consists of species of industrial importance. *C. glutamicum* is used for production of amino acids as well as research. Its gene expression is also greatly regulated by σ factors. Similarly to *Mycobacteria*, this species has one primary σ factor, σ^A, and one primary-like σ factor, σ^B. This regulation is also expanded by five alternative σ factors, σ^C, σ^D, σ^E, σ^H and σ^M. σ^A is responsible for transcription of housekeeping genes and thus is highly expressed during exponential phase, fading when transitioning into stationary phase. The expression of σ^B is the polar opposite and is low during exponential phase and increases when transitioning into stationary phase. This means when the bacterial cell transitions to stationary phase the RNAP holoenzymes contain σ^B rather than σ^A. However, σ^B is a non-essential σ factor. While being able to transcribe some of the housekeeping genes, the presence of σ^B is also associated with stress and periods of slower growth (reviewed in Pátek & Nešvera, 2011).

6.3.1 CoRP RNA

While *Mycobacteria* have Ms1 RNA for regulating gene expression, no homologs were found in *C. glutamicum* (Pfeifer-Sancar et al., 2013; Vaňková Hausnerová et al., 2022). However, using RIP-seq Hausnerová *et al.,* (2024) discovered a new sRNA which can function similarly to the two missing sRNAs. The CoRP RNA (*Corynebacterium* RNAP-binding RNA) transcript was located in an intergenic region which has not been previously annotated. The sRNA is present both during exponential and stationary phase and in total is 516 nt long. Unlike the other mentioned sRNAs, CoRP RNA is cleaved. The predicted structure of the CoRP RNA as well as the cleavage site can be seen in Fig 12. Homologues of CoRP RNA were only found among *Corynebacteria*. Of particular interest is the identified homolog in *C. diphtheriae* (Hausnerová et al., 2024). The potential homolog partially maps to a previously annotated protein, which would mean the sRNA could be translated (Luong et al., 2021).

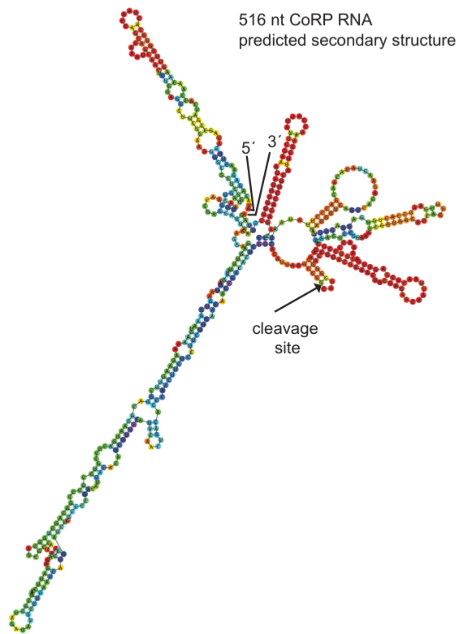

Figure 12. Predicted structure of full-length CoRP RNA in *C. glutamicum*. The figure depicts the predicted secondary structure of the 516 nt-long CoRP RNA. The cleavage site is labelled. The 5´and 3´ends of the sRNA are also labelled. Adapter from (Hausnerová et al., 2024).

While the 516 nt-long RNA is present in low amounts during exponential phase, the CoRP RNA is cleaved during stationary phase. The sRNA, originally 516 nt in length, is cleaved into two RNAs of lengths 318 nt and 198 nt. As seen in Fig 13, the two fragments were labelled the 5´fragment (318 nt fragment) and the 3´ fragment (198 nt fragment). The expression of all the versions of CoRP RNA in both growth phases can also be seen in Fig 13.

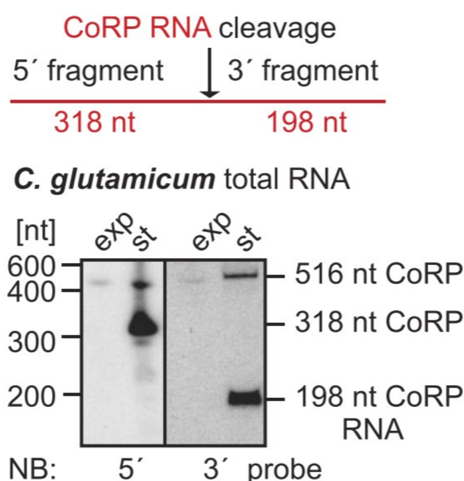

Figure 13. Scheme of CoRP RNA in *C. glutamicum* and northern blot results. On the top of the figure is a scheme of the full length CoRP RNA (length 516 nt). There are the two fragments which are present during stationary phase after their cleavage of the full-length RNA. The two fragments were identified using 5´and 3´ RACE and are labelled accordingly as 5´ fragment and 3´ fragment. The cleavage does not occur in the centre of the RNA and thus a longer and shorter RNA occurs. The longer 5´fragment is 318 nt long while the shorter 3´fragment is 198 nt long. On the bottom of the figure are the northern blot results of the total RNA

detected by a 5´ probe and a 3´probe. The detected bands correspond to the lengths of the fragments determined by 3´and 5´ RACE. According to the visible bands of the northern blot, the full-length RNA was weakly detected during exponential phase using both probes. The full-length RNA was detected strongly in stationary phase using both probes. Using the 5´probe the 5´fragment was strongly detected in stationary phase and using the 3´probe the 3´fragment was strongly detected in stationary phase. The specific bands are labelled in the image with their corresponding lengths to the right of the northern blot image. Adapted from (Hausnerová et al., 2024).

As for the function of CoRP RNA, Hausnerová *et al.* (2024) used RIP-seq (RNA immunoprecipitation sequencing) and found that during both growth phases the full-length (516 nt) RNA was enriched on σ^A, σ^B as well as on the RNAP core. The same results were observed for the 5´and 3´fragment. However, the results of the following glycerol gradient ultracentrifugation suggested CoRP RNA, similarly to Ms1 RNA, also binds the RNAP core without σ factors. Using a protein pull-down experiment to detect which proteins bind to the full-length RNA, we identified RNAP core subunits and no RNAP holoenzymes. The reciprocal approach was used to better characterise the interactions of CoRP RNA and I participated in these experiments. To summarise the findings, the full-length CoRP RNA primarily binds the RNAP core and not the RNAP holoenzyme with σ factors. Complexes with the RNAP holoenzyme thus occur infrequently or are unstable (Hausnerová et al., 2024).

6.4. Bifidobacterium

As introduced to earlier, the genus *Bifidobacteria* includes species which are valuable symbionts important for gut health. Identifying plasmids in these species is not common. However, functional *E. coli* shuttle plasmids for genetic manipulation have been developed (reviewed in Lee & O'Sullivan, 2010). Argnani *et al.,* (1996) developed a protocol to modify *Bifidobacteria*. Using electroporation, they were able to transform several species. They were able to introduce plasmid DNA from a different *Bifidobacterium* species or using plasmid vectors from *Corynebacteria* (Argnani et al., 1996).

6.4.1 Promoters and σ factors

Regulation of gene expression in *Bifidobacteria* is not a deeply investigated topic, as research primarily focuses on health benefits coming from their symbiotic relationship. However, some researchers have analysed and compared their results to what was previously observed in *E. coli*. Details about promoters based on research using *E. coli* were mentioned previously in chapter 2.1.1 on transcription initiation. Using a *Bifidobacterium longum* strain and hidden Markov models, Kozakai *et al.,* (2020) predicted conserved consensus motifs of the -35 and -10 elements. Instead of the expected TTGACA (-35 element) and TATAAT (-10 element) consensus motifs from *E. coli*, they predicted TTGTGC and TACAAT respectively. By observing 37 strains with different constructed plasmids, they concluded the optimal *B. longum* motifs were TTGNNN and TANNNT respectively. "N" in the sequence means any of the four nucleotides (A, C, T, G) can be present. They also observed the spacer between these two elements. In addition to the typical length of 17bp, they identified *Bifidobacteria* also have spacers 11bp long (Kozakai et al., 2020). *B. longum*

possesses at least two σ factors. One of them being a primary σ factor, σ^A. The other is an alternative σ factor, σ^H (Schell et al., 2002). They also compared the *B. longum* σ^A to other primary σ factors from *E. coli* and *Bacillus subtilis*. The N-terminus of σ^A showed a well-defined polar domain which consisted of a positively and negatively charged domain. The polar domain includes 129 amino acids in total. The domain is conserved among *Bifidobacteria* as well as the *Actinobacteria* phylum. The domain could potentially influence promoter recognition and thus the structure of promoter sequences (Kozakai et al., 2020).

6.4.2 WhiB-like family proteins

Many species of *Bifidobacteria* encode WhiB-like family proteins (Averina et al., 2012). The family of WhiB-like proteins is specific to *Actinobacteria,* including the genera *Mycobacteria*, *Corynebacteria*, *Streptomyces* and *Bifidobacteria*. There were over 270 homologs of WhiB-like genes identified (Gao et al., 2006). The WhiB gene was first identified in *S. coelicolor* in 1972 and is essential for sporulation in this species (Chater, 1972). WhiB3 in *M. tuberculosis* affects the expression from several promoters which influence the immune response to the infection (Steyn et al., 2002). In general, WhiB-like family proteins are small proteins which act as transcription regulators, regulating gene expression in response to various stresses (Geiman et al., 2006). The study by Averina *et al.,* (2012) identified and characterised the WhiB-like (Wbl) family proteins in 36 strains from 11 species of *Bifidobacteria*. The bioinformatically-identified genes were divided into two categories. Some encoded WhiB2 proteins consisting of 99-117 amino acid residues while others encoded WblE proteins consisting of 73-96 amino acid residues (Averina et al., 2012).

As for the structure, the identified bifidobacterial Wbl proteins have four conserved cysteine residues within the N-terminus and a putative helix-turn-helix motif at the C-terminus. In Wbl proteins, the conserved cysteine residues typically bind a [4Fe-4S] cluster (Jakimowicz et al., 2005). As part of the helix-turn helix motif, there is a conserved G(V/I)WGGLSE motif in bifidobacterial Wbl proteins (Averina et al., 2012). The C-terminus was previously proposed to have a DNA-binding motif (Soliveri et al., 2000). The first available structure of a Wbl protein was obtained in 2017 and can be seen in Fig 14. It is the structure of WhiB1 from *M. tuberculosis*. The NMR model revealed WhiB1 has four helices. Three core α-helices are held together by a [4Fe-4S] cluster. The fourth helix was identified as the C-terminal helix. Apart from these findings, the protein appeared to be disordered (Kudhair et al., 2017). WhiB1 was previously identified to interact with the primary σ factor, σ^A (Feng et al., 2016). Later research showed that interacting with σ^A was dependent on the presence of the [4Fe-4S] cluster (Kudhair et al., 2017).

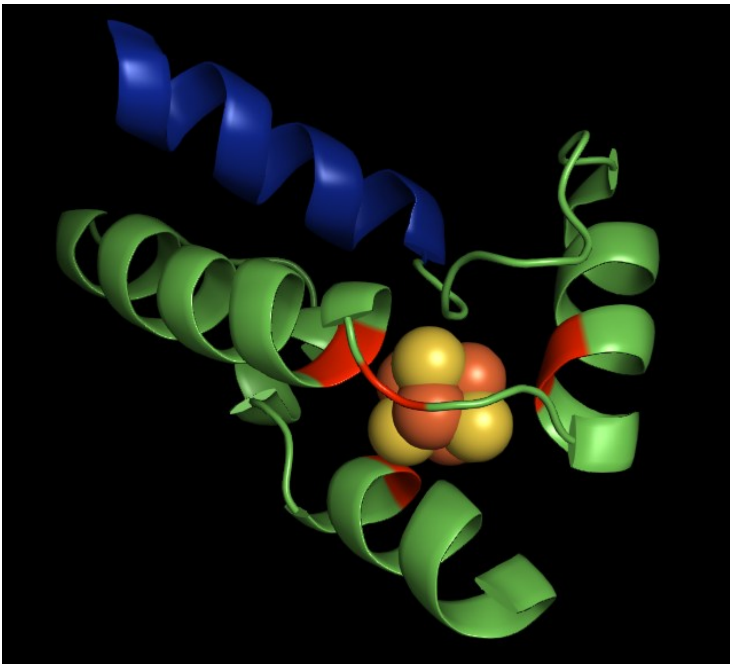

Figure 14. Structure model of *M. tuberculosis* WhiB1. The figure depicts the secondary structure of WhiB1. The model shows amino acid residues 1-76. Residues 77-84 are not visible as their structure is disordered. In green are the three core α-helices with the [4Fe-4S] cluster in the middle (orange and yellow atoms). The four conserved cysteine residues are coloured red (C9, C37, C40 and C46) and are located within the three core α-helices. In blue is the C-terminal helix. Apart from the four helices, there is a short helical segment between two of the core α-helices (Kudhair et al., 2017). The figure shows the PDB structure according to Kudhair *et al.,* (2017). Using PyMOL, I colour-coded the PDB structure (Williamson et al., 2017) to illustrate the previously described elements.

The study by Averina *et al.,* (2012) identified and characterised the WhiB-like (Wbl) family proteins in 36 strains from 11 species of *Bifidobacteria*. The bioinformatically-identified genes were divided into two categories. Some encoded WhiB2 proteins consisting of 99-117 amino acid residues while others encoded WblE proteins consisting of 73-96 amino acid residues (Averina et al., 2012). Using *B. longum* they observed mRNA levels of WhiB2 and WblE genes during exponential and stationary phase. While the levels of WhiB2 were constant between both phases, the levels of WblE increased when progressing into stationary phase. They also observed expression under different stresses. Under the tested stresses WhiB2 levels remained constant or decreased while WblE levels increased. An increase in WblE levels was observed during starvation, osmotic stress, oxidative stress and treatments with tetracycline and bile-salts. Alignment of the amino acid sequences showed the WhiB2 from *B. longum* is similar to WhiB2 from *M. tuberculosis*. In case of WblE, it is similar to WhiB1 from *M. tuberculosis* and WblE from *S. coelicolor* (Averina et al., 2012).

7. Conclusion

In this Thesis, I compared transcription regulation in various species of *Actinobacteria*. This overview focused on regulations by proteins and regulatory RNAs found in *Actinobacteria*. Precisely, what role do these regulations have in the function of RNAP and σ factors as part of the transcription machinery. As for the σ factors, the focus was on the primary σ^A and primary-like σ^B. The regulatory RNAs included Ms1 RNA and CoRP RNA. Furthermore, this Thesis closely explored regulations in the genera *Mycobacterium*, *Streptomyces*, *Corynebacterium* and *Bifidobacterium*. While most of the characterised regulators are specific to *Actinobacteria*, their homologues can be often found beyond this phylum. Thus, the distribution of these regulators among the domain *Bacteria* was mentioned as well.

The RbpA protein was found to regulate transcription by stabilising the RPo by interacting with the RNAP holoenzyme. To interact, RNAP holoenzyme can contains only the primary and primary-like σ factors, σ^A and σ^B. RbpA is considered a transcription activator important during stationary phase and stress responses. The protein is also associated with resistance to rifampicin. RbpA is specific to *Actinobacteria*, its role described in *Mycobacteria* as well as in *Streptomyces*.

The CarD protein is an essential regulator in *Mycobacteria*, stabilising the RPo and regulating rRNA transcription. It binds the RNAP holoenzyme and is considered an rRNA transcription activator. The role of CarD during starvation was long debated, but current data suggests CarD protein levels decrease in response to starvation to decrease the metabolic rate of the cell. While CarD has homologues across many phyla, it is not found in *E. coli*. In *M. tuberculosis* it is key for tuberculosis persistence, and it is also highly conserved in other *Mycobacteria*.

HelD proteins are helicase-like factors, widespread among many phyla and thus divided into three classes. The class II HelD present in *Mycobacteria* is a transcription factor responsible for recycling RNAP. HelD binds to the RNAP core and clears out any nucleic acids, to free stalled RNAP and RNAP which has not dissociated after termination. On the contrary, a complex of HelD with a σ^A-containing RNAP holoenzyme and RbpA was also identified. HelD is also associated with resistance to rifampicin in *M. smegmatis* or *M. abscessus*.

The protein MoaB2 interacts with σ^A, inhibiting transcription dependent on σ^A, the primary σ factor. Sequestering of σ^A limits the formation of the σ^A-containing RNAP holoenzyme necessary for transcribing housekeeping genes. MoaB2 is present both during exponential and stationary phase and only binds σ^A when it is not in complex with the RNAP core. It does not sequester alternative σ factors, including primary-like σ^B. MoaB2 is present in *M. smegmatis* and *M. marinum,* but is not an essential transcription factor. It also has two homologs in *E. coli.*

Transcription regulations by sRNAs in the genera *Mycobacterium*, *Streptomyces* and *Corynebacterium* were characterised. Ms1 RNA identified in *M. smegmatis* binds to the RNAP core. It does not interact with the σ^A-RNAP holoenzyme and is expressed mainly during stationary phase. It has the secondary structure of a long hairpin with a central bubble important for binding the RNAP core. During stationary phase the stability of Ms1 greatly increases in comparison to exponential phase. Depletion of Ms1 RNA leads to lower levels of RNAP, explaining why Ms1 RNA is associated with accelerated outgrowth.

However, Ms1 is not essential for viable growth. Ms1 RNA has many homologues; MST2823 in *M. tuberculosis* and scr3559 in *S. coelicolor* in addition to other homologues identified in 824 species of *Actinobacteria*. No Ms1 homologues were found in *C. glutamicum* as *Corynebacteria* appear to utilise a different sRNA, the CoRP RNA. CoRP RNA has a similar structure to Ms1 RNA and is present both in exponential and stationary phase. Similarly to Ms1 RNA, it binds the RNAP core. During stationary phase it is cleaved into two fragments.

WhiB-like family proteins specific to *Actinobacteria* are encoded by many species of *Bifidobacteria*. These transcription factors regulate gene expression in response to stresses. WhiB-like family proteins were identified in 11 species of *Bifidobacteria* and divided into two categories: WhiB2 and WblE proteins. WblE proteins increase during stationary phase and under various stresses.

To conclude, this overview compared transcription regulations in various species and genera of *Actinobacteria*. It can thus help identify areas of interest for future research. Like CarD with CrsL, these regulators may have other important interacting partners which remain unknown. This illustrates, how transcriptional regulations in *Actinobacteria* leave much space for further research.

8. References

Secondary citation *

Argnani, A., Leer, R. J., van Luijk, N., & Pouwels, P. H. (1996). A convenient and reproducible method to genetically transform bacteria of the genus *Bifidobacterium*. *Microbiology*, *142*(1), 109–114. https://doi.org/10.1099/13500872-142-1-109

Averina, O. V., Zakharevich, N. V., & Danilenko, V. N. (2012). Identification and characterization of WhiB-like family proteins of the *Bifidobacterium* genus. *Anaerobe*, *18*(4), 421–429. https://doi.org/10.1016/j.anaerobe.2012.04.011

Aziz, D. B., Low, J. L., Wu, M.-L., Gengenbacher, M., Teo, J. W. P., Dartois, V., & Dick, T. (2017). Rifabutin Is Active against *Mycobacterium abscessus* Complex. *Antimicrobial Agents and Chemotherapy*, *61*(6). https://doi.org/10.1128/AAC.00155-17

Bae, B., Chen, J., Davis, E., Leon, K., Darst, S. A., & Campbell, E. A. (2015). CarD uses a minor groove wedge mechanism to stabilize the RNA polymerase open promoter complex. *ELife*, *4*. https://doi.org/10.7554/eLife.08505

* Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y., & van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of *Actinobacteria*. *Microbiology and Molecular Biology Reviews*, *80*(1), 1–43. https://doi.org/10.1128/MMBR.00019-15

Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L., & Breaker, R. R. (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA*, *11*(5), 774–784. https://doi.org/10.1261/rna.7286705

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., … Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, *417*(6885), 141–147. https://doi.org/10.1038/417141a

Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A., & Duncan, K. (2002). Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Molecular Microbiology*, *43*(3), 717–731. https://doi.org/10.1046/j.1365-2958.2002.02779.x

Bevers, L. E., Hagedoorn, P.-L., Santamaria-Araujo, J. A., Magalon, A., Hagen, W. R., & Schwarz, G. (2008). Function of MoaB Proteins in the Biosynthesis of the Molybdenum and Tungsten Cofactors. *Biochemistry*, *47*(3), 949–956. https://doi.org/10.1021/bi7020487

Bevilacqua, L., Ovidi, M., Di Mattia, E., Trovatelli, L. D., & Canganella, F. (2003). Screening of *Bifidobacterium* strains isolated from human faeces for antagonistic activities against potentially bacterial pathogens. *Microbiological Research*, *158*(2), 179–185. https://doi.org/10.1078/0944-5013-00192

Blatter, E. E., Ross, W., Tang, H., Gourse, R. L., & Ebright, R. H. (1994). Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell*, *78*(5), 889–896. https://doi.org/10.1016/S0092-8674(94)90682-3

Bobek, J., Mikulová, A., Šetinová, D., Elliot, M., & Čihák, M. (2021). 6S-Like scr3559 RNA Affects Development and Antibiotic Production in *Streptomyces coelicolor*. *Microorganisms*, *9*(10), 2004. https://doi.org/10.3390/microorganisms9102004

Boyaci, H., Chen, J., Lilic, M., Palka, M., Mooney, R. A., Landick, R., Darst, S. A., & Campbell, E. A. (2018). Fidaxomicin jams *Mycobacterium tuberculosis* RNA polymerase motions needed for initiation via RbpA contacts. *ELife*, *7*. https://doi.org/10.7554/eLife.34823

Brezovská, B., Narasimhan, S., Šiková, M., Šanderová, H., Kovaľ, T., Borah, N., Shoman, M., Pospíšilová, D., Vaňková Hausnerová, V., Tužinčin, D., Černý, M., Komárek, J., Janoušková, M., Kambová, M., Halada, P., Křenková, A., Hubálek, M., Trundová, M., Dohnálek, J., … Krásný, L. (2024). MoaB2, a newly identified transcription factor, binds to σ^{A} in *Mycobacterium smegmatis*. *Journal of Bacteriology*. https://doi.org/10.1128/jb.00066-24

*Browning, D. F., & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, *2*(1), 57–65. https://doi.org/10.1038/nrmicro787

Buchanan, R. E. (1918). Life Phases in a Bacterial Culture. *Source: The Journal of Infectious Diseases*, *23*(2), 109–125. https://about.jstor.org/terms

Burgess, R. R., Travers, A. A., Dunn, J. J., & Bautz, E. K. F. (1969). Factor Stimulating Transcription by RNA Polymerase. *Nature*, *221*, 43–46.

Buttner, M. J., Chater, K. F., & Bibb, M. J. (1990). Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, *172*(6), 3367–3378. https://doi.org/10.1128/jb.172.6.3367-3378.1990

Buttner, M. J., & Lewis, C. G. (1992). Construction and characterization of *Streptomyces coelicolor* A3(2) mutants that are multiply deficient in the nonessential hrd-encoded RNA polymerase sigma factors. *Journal of Bacteriology*, *174*(15), 5165–5167. https://doi.org/10.1128/jb.174.15.5165-5167.1992

Callaham, D., Del Tredici, P., & Torrey, J. G. (1978). Isolation and Cultivation in vitro of the Actinomycete Causing Root Nodulation in *Comptonia*. *Science*, *199*(4331), 899–902. https://doi.org/10.1126/science.199.4331.899

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell*, *104*(6), 901–912. https://doi.org/10.1016/S0092-8674(01)00286-0

Carafa, Y. d'Aubenton, Brody, E., & Thermes, C. (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. *Journal of Molecular Biology*, *216*(4), 835–858. https://doi.org/10.1016/S0022-2836(99)80005-9

Carpousis, A. J., & Gralla, J. D. (1980). Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry*, *19*(14), 3245–3253. https://doi.org/10.1021/bi00555a023

Carrasco, B., Fernández, S., Petit, M.-A., & Alonso, J. C. (2001). Genetic Recombination in *Bacillus subtilis* 168: Effect of Δ *helD* on DNA Repair and Homologous Recombination. *Journal of Bacteriology*, *183*(19), 5772–5777. https://doi.org/10.1128/JB.183.19.5772-5777.2001

Celmer, W. D., Chmurny, G. N., Moppett, C. E., Ware, R. S., Watts, P. C., & Whipple, E. B. (1980). Structure of natural antibiotic CP-47,444. *Journal of the American Chemical Society*, *102*(12), 4203–4209. https://doi.org/10.1021/ja00532a036

Chater, K. F. (1972). A Morphological and Genetic Mapping Study of White Colony Mutants of *Streptomyces coelicolor*. *Journal of General Microbiology*, *72*(1), 9–28. https://doi.org/10.1099/00221287-72-1-9

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Davis, E., Chen, J., Leon, K., Darst, S. A., & Campbell, E. A. (2015). Mycobacterial RNA polymerase forms unstable open promoter complexes that are stabilized by CarD. *Nucleic Acids Research*, *43*(1), 433–445. https://doi.org/10.1093/nar/gku1231

Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savery, N. J., & Darst, S. A. (2006). Structural Basis for Bacterial Transcription-Coupled DNA Repair. *Cell*, *124*(3), 507–520. https://doi.org/10.1016/j.cell.2005.11.045

Delumeau, O., Lecointe, F., Muntel, J., Guillot, A., Guédon, E., Monnet, V., Hecker, M., Becher, D., Polard, P., & Noirot, P. (2011). The dynamic protein partnership of RNA polymerase in *Bacillus subtilis*. *PROTEOMICS*, *11*(15), 2992–3001. https://doi.org/10.1002/pmic.201000790

Dey, A., Adithi, V. R., & Chatterji, D. (2012). Co-evolution of RNA polymerase with RbpA in the phylum *Actinobacteria*. *Applied & Translational Genomics*, *1*, 9–20. https://doi.org/10.1016/j.atg.2012.03.001

Dey, A., Verma, A. K., & Chatterji, D. (2010). Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. *Microbiology*, *156*(3), 873–883. https://doi.org/10.1099/mic.0.033670-0

Dickson, R. C., Abelson, J., Barnes, W. M., & Reznikoff, W. S. (1975). Genetic Regulation: The Lac Control Region. *Science*, *187*(4171), 27–35. https://doi.org/10.1126/science.1088926

Dombroski, A. J., Walter, W. A., & Gross, C. A. (1993). The role of the sigma subunit in promoter recognition by RNA polymerase. *Cellular & Molecular Biology Research*, *39*(4), 311–317.

Farnham, P. J., & Platt, T. (1980). A model for transcription termination suggested by studies on the trp attenuator in vitro using base analogs. *Cell*, *20*(3), 739–748. https://doi.org/10.1016/0092-8674(80)90320-7

* Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial sigma factors: A historical, structural, and genomic perspective. *Annual Review of Microbiology*, *68*, 357–376. https://doi.org/10.1146/annurev-micro-092412-155737

Feng, L., Chen, Z., Wang, Z., Hu, Y., & Chen, S. (2016). Genome-wide characterization of monomeric transcriptional regulators in *Mycobacterium tuberculosis*. *Microbiology*, *162*(5), 889–897. https://doi.org/10.1099/mic.0.000257

Fredrick, K., & Helmann, J. D. (1997). RNA polymerase sigma factor determines start-site selection but is not required for upstream promoter element activation on heteroduplex (bubble) templates. *Proceedings of the National Academy of Sciences*, *94*(10), 4982–4987. https://doi.org/10.1073/pnas.94.10.4982

Gande, R., Gibson, K. J. C., Brown, A. K., Krumbach, K., Dover, L. G., Sahm, H., Shioyama, S., Oikawa, T., Besra, G. S., & Eggeling, L. (2004). Acyl-CoA Carboxylases (accD2 and accD3), Together with a Unique Polyketide Synthase (Cg-pks), Are Key to Mycolic Acid Biosynthesis in *Corynebacterianeae* Such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis*. *Journal of Biological Chemistry*, *279*(43), 44847–44857. https://doi.org/10.1074/jbc.M408648200

Gao, B., Paramanathan, R., & Gupta, R. S. (2006). Signature proteins that are distinctive characteristics of *Actinobacteria* and their subgroups. *Antonie van Leeuwenhoek*, *90*(1), 69–91. https://doi.org/10.1007/s10482-006-9061-2

Garner, A. L., Rammohan, J., Huynh, J. P., Onder, L. M., Chen, J., Bae, B., Jensen, D., Weiss, L. A., Manzano, A. R., Darst, S. A., Campbell, E. A., Nickels, B. E., Galburt, E. A., & Stallings, C. L. (2017). Effects of Increasing the Affinity of CarD for RNA Polymerase on *Mycobacterium tuberculosis* Growth, rRNA Transcription, and Virulence. *Journal of Bacteriology*, *199*(4). https://doi.org/10.1128/JB.00698-16

Geiman, D. E., Raghunand, T. R., Agarwal, N., & Bishai, W. R. (2006). Differential Gene Expression in Response to Exposure to Antimycobacterial Agents and Other Stress Conditions among Seven *Mycobacterium tuberculosis whiB-* Like Genes. *Antimicrobial Agents and Chemotherapy*, *50*(8), 2836–2841. https://doi.org/10.1128/AAC.00295-06

Gomez, M., Doukhan, L., Nair, G., & Smith, I. (1998). *sigA* is an essential gene in *Mycobacterium smegmatis*. *Molecular Microbiology*, *29*(2), 617–628. https://doi.org/10.1046/j.1365-2958.1998.00960.x

Gulten, G., & Sacchettini, J. C. (2013). Structure of the Mtb CarD/RNAP β-Lobes Complex Reveals the Molecular Basis of Interaction and Presents a Distinct DNA-Binding Domain for *Mtb* CarD. *Structure*, *21*(10), 1859–1869. https://doi.org/10.1016/j.str.2013.08.014

Hausnerová, V. V., Shoman, M., Kumar, D., Schwarz, M., Modrák, M., Matějčková, J. J., Mikesková, E., Neva, S., Herrmannová, A., Šiková, M., Halada, P., Novotná, I., Pajer, P., Valášek, L. S., Převorovský, M., Krásný, L., & Hnilicová, J. (2024). RIP-seq reveals RNAs that interact with RNA polymerase and primary sigma factors in bacteria. *Nucleic Acids Research*, *52*(8), 4604–4626. https://doi.org/10.1093/nar/gkae081

Hengge-Aronis, R. (1993). Survival of hunger and stress: The role of rpoS in early stationary phase gene regulation in *E. coli*. *Cell*, *72*(2), 165–168. https://doi.org/10.1016/0092-8674(93)90655-A

Hindley, J. (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *Journal of Molecular Biology*, *30*(1), 125–136. https://doi.org/10.1016/0022-2836(67)90248-3

Hnilicová, J., Jirát Matějčková, J., Šiková, M., Pospíšil, J., Halada, P., Pánek, J., & Krásný, L. (2014). Ms1, a novel sRNA interacting with the RNA polymerase core in *mycobacteria*. *Nucleic Acids Research*, *42*(18), 11763–11776. https://doi.org/10.1093/nar/gku793

Hook-Barnard, I. G., & Hinton, D. M. (2009). The promoter spacer influences transcription initiation via σ^{70} region 1.1 of *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences*, *106*(3), 737–742. https://doi.org/10.1073/pnas.0808133106

* Hsu, L. M. (2002). Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, *1577*(2), 191–207. https://doi.org/10.1016/S0167-4781(02)00452-9

Hu, Y., Morichaud, Z., Chen, S., Leonetti, J. P., & Brodolin, K. (2012a). *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ a-containing RNA polymerase holoenzyme. *Nucleic Acids Research*, *40*(14), 6547–6557. https://doi.org/10.1093/nar/gks346

Hu, Y., Morichaud, Z., Chen, S., Leonetti, J.-P., & Brodolin, K. (2012b). *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ A -containing RNA polymerase holoenzyme. *Nucleic Acids Research*, *40*(14), 6547–6557. https://doi.org/10.1093/nar/gks346

Hu, Y., Morichaud, Z., Sudalaiyadum Perumal, A., Roquet-Baneres, F., & Brodolin, K. (2014). *Mycobacterium* RbpA cooperates with the stress-response σB subunit of RNA polymerase in promoter DNA unwinding. *Nucleic Acids Research*, *42*(16), 10399–10408. https://doi.org/10.1093/nar/gku742

Hubin, E. A., Fay, A., Xu, C., Bean, J. M., Saecker, R. M., Glickman, M. S., Darst, S. A., & Campbell, E. A. (2017). Structure and function of the mycobacterial transcription initiation complex with the essential regulator RbpA. *ELife*, *6*. https://doi.org/10.7554/eLife.22520

Hurst-Hess, K., Biswas, R., Yang, Y., Rudra, P., Lasek-Nesselquist, E., & Ghosh, P. (2019). Mycobacterial SigA and SigB Cotranscribe Essential Housekeeping Genes during Exponential Growth. *MBio*, *10*(3). https://doi.org/10.1128/mBio.00273-19

Hurst-Hess, K. R., Saxena, A., Rudra, P., Yang, Y., & Ghosh, P. (2022). *Mycobacterium abscessus* HelR interacts with RNA polymerase to confer intrinsic rifamycin resistance. *Molecular Cell*, *82*(17), 3166-3177.e5. https://doi.org/10.1016/j.molcel.2022.06.034

Jakimowicz, P., Cheesman, M. R., Bishai, W. R., Chater, K. F., Thomson, A. J., & Buttner, M. J. (2005). Evidence That the Streptomyces Developmental Protein WhiD, a Member of the WhiB Family, Binds a [4Fe-4S] Cluster. *Journal of Biological Chemistry*, *280*(9), 8309–8315. https://doi.org/10.1074/jbc.M412622200

Kouba, T., Koval', T., Sudzinová, P., Pospíšil, J., Brezovská, B., Hnilicová, J., Šanderová, H., Janoušková, M., Šiková, M., Halada, P., Sýkora, M., Barvík, I., Nováček, J., Trundová, M., Dušková, J., Skálová, T., Chon, Ur., Murakami, K. S., Dohnálek, J., & Krásný, L. (2020). Mycobacterial HelD is a nucleic acids-clearing factor for RNA polymerase. *Nature Communications*, *11*(1), 6419. https://doi.org/10.1038/s41467-020-20158-4

Kovaľ, T., Borah, N., Sudzinová, P., Brezovská, B., Šanderová, H., Vaňková Hausnerová, V., Křenková, A., Hubálek, M., Trundová, M., Adámková, K., Dušková, J., Schwarz, M., Wiedermannová, J., Dohnálek, J., Krásný, L., & Kouba, T. (2024). Mycobacterial HelD connects RNA polymerase recycling with

transcription initiation. *Nature Communications*, *15*(1), 8740. https://doi.org/10.1038/s41467-024-52891-5

Kozakai, T., Izumi, A., Horigome, A., Odamaki, T., Xiao, J., Nomura, I., & Suzuki, T. (2020). Structure of a Core Promoter in *Bifidobacterium longum* NCC2705. *Journal of Bacteriology*, *202*(7). https://doi.org/10.1128/JB.00540-19

Kozmin, S. G., & Schaaper, R. M. (2013). Genetic characterization of moaB mutants of *Escherichia coli*. *Research in Microbiology*, *164*(7), 689–694. https://doi.org/10.1016/j.resmic.2013.05.001

Kudhair, B. K., Hounslow, A. M., Rolfe, M. D., Crack, J. C., Hunt, D. M., Buxton, R. S., Smith, L. J., Le Brun, N. E., Williamson, M. P., & Green, J. (2017). Structure of a Wbl protein and implications for NO sensing by *M. tuberculosis*. *Nature Communications*, *8*(1), 2280. https://doi.org/10.1038/s41467-017-02418-y

Lane, W. J., & Darst, S. A. (2010). Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis. *Journal of Molecular Biology*, *395*(4), 671–685. https://doi.org/10.1016/j.jmb.2009.10.062

Larsen, J. S., Miller, M., Oakley, A. J., Dixon, N. E., & Lewis, P. J. (2021). Multiple classes and isoforms of the RNA polymerase recycling motor protein HelD. *MicrobiologyOpen*, *10*(6), e1251. https://doi.org/10.1002/mbo3.1251

* Lee, J.-H., & O'Sullivan, D. J. (2010). Genomic Insights into *Bifidobacteria*. *Microbiology and Molecular Biology Reviews*, *74*(3), 378–416. https://doi.org/10.1128/MMBR.00004-10

Leibman, M., & Hochschild, A. (2007). A σ-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *The EMBO Journal*, *26*(6), 1579–1590. https://doi.org/10.1038/sj.emboj.7601612

Li, X., Chen, F., Liu, X., Xiao, J., Andongma, B. T., Tang, Q., Cao, X., Chou, S.-H., Galperin, M. Y., & He, J. (2022). Clp protease and antisense RNA jointly regulate the global regulator CarD to mediate mycobacterial starvation response. *ELife*, *11*. https://doi.org/10.7554/eLife.73347

Lonetto, M., Gribskov, M., & Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology*, *174*(12), 3843–3849. https://doi.org/10.1128/jb.174.12.3843-3849.1992

Luong, T. T., Nguyen, M. T., Chen, Y.-W., Chang, C., Lee, J. H., Wittchen, M., Ton-That, H., Cruz, M., Garsin, D. A., Das, A., Tauch, A., & Ton-That, H. (2021). Ribonuclease J-Mediated mRNA Turnover Modulates Cell Shape, Metabolism and Virulence in *Corynebacterium diphtheriae*. *Microorganisms*, *9*(2), 389. https://doi.org/10.3390/microorganisms9020389

Martin, F. H., & Tinoco, I. (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Research*, *8*(10), 2295–2300. https://doi.org/10.1093/nar/8.10.2295

Merrick, M. J. (1993). In a class of its own — the RNA polymerase sigma factor σ; ^{54} (σ ^{N}). *Molecular Microbiology*, *10*(5), 903–909. https://doi.org/10.1111/j.1365-2958.1993.tb00961.x

Mikulík, K., Bobek, J., Zídková, J., & Felsberg, J. (2014). 6S RNA modulates growth and antibiotic production in *Streptomyces coelicolor*. *Applied Microbiology and Biotechnology*, *98*(16), 7185–7197. https://doi.org/10.1007/s00253-014-5806-4

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H., & Severinov, K. (2001). Bacterial RNA polymerase omega subunit and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *PNAS*, *98*(3), 892–897. www.rcsb.org

Minnikin, D. E., & Polgar, N. (1966). Studies on the mycolic acids from human tubercle bacilli. *Tetrahedron Letters*, *7*(23), 2643–2647. https://doi.org/10.1016/S0040-4039(01)84131-9

Mukherjee, K., & Chatterji, D. (1997). Studies on the ω Subunit of *Escherichia Coli* RNA Polymerase. *European Journal of Biochemistry*, *247*(3), 884–889. https://doi.org/10.1111/j.1432-1033.1997.00884.x

Murakami, K. S. (2013). X-ray Crystal Structure of *Escherichia coli* RNA Polymerase σ70 Holoenzyme. *Journal of Biological Chemistry*, *288*(13), 9126–9134. https://doi.org/10.1074/jbc.M112.430900

Newell, K. V., Thomas, D. P., Brekasis, D., & Paget, M. S. B. (2006). The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. *Molecular Microbiology*, *60*(3), 687–696. https://doi.org/10.1111/j.1365-2958.2006.05116.x

Newing, T. P., Oakley, A. J., Miller, M., Dawson, C. J., Brown, S. H. J., Bouwer, J. C., Tolun, G., & Lewis, P. J. (2020). Molecular basis for RNA polymerase-dependent transcription complex recycling by the helicase-like motor protein HelD. *Nature Communications*, *11*(1), 6420. https://doi.org/10.1038/s41467-020-20157-5

Nichols, J., & Rajagopalan, K. V. (2002). *Escherichia coli* MoeA and MogA. *Journal of Biological Chemistry*, *277*(28), 24995–25000. https://doi.org/10.1074/jbc.M203238200

Nicolás, F. J., Ruiz-Vázquez, R. M., & Murillo, F. J. (1994). A genetic link between light response and multicellular development in the bacterium *Myxococcus xanthus*. *Genes & Development*, *8*(19), 2375–2387. https://doi.org/10.1101/gad.8.19.2375

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., Pukall, R., Klenk, H.-P., Goodfellow, M., & Göker, M. (2018). Genome-Based Taxonomic Classification of the Phylum *Actinobacteria*. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.02007

* Paget, M. S. (2015). Bacterial sigma factors and anti-sigma factors: Structure, function and distribution. *Biomolecules*, *5*(3), 1245–1265. https://doi.org/10.3390/biom5031245

Paget, M. S. B., Kang J. G., Roe J.H., & Buttner M.J. (1998). sigma R, an RNA polymerase sigma factor that modulates expression of the thioredoxin system in response to oxidative stress in *Streptomyces coelicolor* A3(2). *The EMBO Journal*, *17*(19), 5776–5782. https://doi.org/10.1093/emboj/17.19.5776

Paget, M. S. B., Molle, V., Cohen, G., Aharonowitz, Y., & Buttner, M. J. (2001). Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the σ^{R} regulon. *Molecular Microbiology*, *42*(4), 1007–1020. https://doi.org/10.1046/j.1365-2958.2001.02675.x

* Paget, M. S., & Helmann, J. D. (2003). The σ70family of sigma factors. *Genome Biology*, *4*(1), 203. https://doi.org/10.1186/gb-2003-4-1-203

Panek, J., Krasny, L., Bobek, J., Jezkova, E., Korelusova, J., & Vohradsky, J. (2011). The suboptimal structures find the optimal RNAs: homology search for bacterial non-coding RNAs using suboptimal RNA structures. *Nucleic Acids Research*, *39*(8), 3418–3426. https://doi.org/10.1093/nar/gkq1186

* Pátek, M., & Nešvera, J. (2011). Sigma factors and promoters in *Corynebacterium glutamicum*. *Journal of Biotechnology*, *154*(2–3), 101–113. https://doi.org/10.1016/j.jbiotec.2011.01.017

Paul, B. J., Barker, M. M., Ross, W., Schneider, D. A., Webb, C., Foster, J. W., & Gourse, R. L. (2004). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell*, *118*(3), 311–322. https://doi.org/10.1016/j.cell.2004.07.009

Perederina, A., Svetlov, V., Vassylyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I., & Vassylyev, D. G. (2004). Regulation through the Secondary Channel—Structural Framework for ppGpp-DksA Synergism during Transcription. *Cell*, *118*(3), 297–309. https://doi.org/10.1016/j.cell.2004.06.030

Pfeifer-Sancar, K., Mentz, A., Rückert, C., & Kalinowski, J. (2013). Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics*, *14*(1), 888. https://doi.org/10.1186/1471-2164-14-888

Prusa, J., Jensen, D., Santiago-Collazo, G., Pope, S. S., Garner, A. L., Miller, J. J., Ruiz Manzano, A., Galburt, E. A., & Stallings, C. L. (2018). Domains within RbpA Serve Specific Functional Roles That Regulate the Expression of Distinct Mycobacterial Gene Subsets. *Journal of Bacteriology*, *200*(13). https://doi.org/10.1128/JB.00690-17

Prusa, J., Zhu, D. X., Flynn, A. J., Jensen, D., Ruiz Manzano, A., Galburt, E. A., & Stallings, C. L. (2022). Molecular dissection of RbpA-mediated regulation of fidaxomicin sensitivity in *mycobacteria*. *Journal of Biological Chemistry*, *298*(4), 101752. https://doi.org/10.1016/j.jbc.2022.101752

Raghavan, A., & Chatterji, D. (1998). Guanosine tetraphosphate-induced dissociation of open complexes at the *Escherichia coli* ribosomal protein promoters rplJ and rpsA P1: nanosecond depolarization spectroscopic studies. *Biophysical Chemistry*, *75*(1), 21–32. https://doi.org/10.1016/S0301-4622(98)00186-0

Rammohan, J., Ruiz Manzano, A., Garner, A. L., Stallings, C. L., & Galburt, E. A. (2015). CarD stabilizes mycobacterial open complexes via a two-tiered kinetic mechanism. *Nucleic Acids Research*, *43*(6), 3272–3285. https://doi.org/10.1093/nar/gkv078

Rawat, M., Johnson, C., Cadiz, V., & Av-Gay, Y. (2007). Comparative analysis of mutants in the mycothiol biosynthesis pathway in *Mycobacterium smegmatis*. *Biochemical and Biophysical Research Communications*, *363*(1), 71–76. https://doi.org/10.1016/j.bbrc.2007.08.142

* Rebets, Y., Tsolis, K. C., Guðmundsdóttir, E. E., Koepff, J., Wawiernia, B., Busche, T., Bleidt, A., Horbal, L., Myronovskyi, M., Ahmed, Y., Wiechert, W., Rückert, C., Hamed, M. B., Bilyk, B., Anné, J., Friðjónsson, Ó., Kalinowski, J., Oldiges, M., Economou, A., & Luzhetskyy, A. (2018). Characterization of Sigma Factor Genes in *Streptomyces lividans* TK24 Using a Genomic Library-Based Approach for Multiple Gene Deletions. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.03033

Rhodius, V. A., Segall-Shapiro, T. H., Sharon, B. D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D. H., Clancy, K., Peterson, T. C., Gross, C. A., & Voigt, C. A. (2013). Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Molecular Systems Biology*, *9*(1). https://doi.org/10.1038/msb.2013.58

Roberts, J. W. (1969). Termination Factor for RNA Synthesis. *Nature*, *224*(5225), 1168–1174. https://doi.org/10.1038/2241168a0

Rominski, A., Roditscheff, A., Selchow, P., Böttger, E. C., & Sander, P. (2017). Intrinsic rifamycin resistance of *Mycobacterium abscessus* is mediated by ADP-ribosyltransferase MAB_0591. *Journal of Antimicrobial Chemotherapy*, *72*(2), 376–384. https://doi.org/10.1093/jac/dkw466

* Ruiz, N., & Silhavy, T. J. (2022). How *Escherichia coli* Became the Flagship Bacterium of Molecular Biology. *Journal of Bacteriology*, *204*(9). https://doi.org/10.1128/jb.00230-22

* Santangelo, T. J., & Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. *Nature Reviews Microbiology*, *9*(5), 319–329. https://doi.org/10.1038/nrmicro2560

Schell, M. A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M.-C., Desiere, F., Bork, P., Delley, M., Pridmore, R. D., & Arigoni, F. (2002). The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proceedings of the National Academy of Sciences*, *99*(22), 14422–14427. https://doi.org/10.1073/pnas.212527599

Shoman, M., Jirát Matějčková, J., Schwarz, M., Černý, M., Borah, N., Vaňková Hausnerová, V., Šiková, M., Šanderová, H., Halada, P., Hubálek, M., Dvořáková, V., Převorovský, M., Holubová, J., Staněk, O., Krásný, L., Žídek, L., & Hnilicová, J. (2024). Expanding the CarD interaction network: CrsL is a novel transcription factor in *Mycobacterium smegmatis*. https://doi.org/10.1101/2024.12.02.626298

Šiková, M., Janoušková, M., Ramaniuk, O., Páleníková, P., Pospíšil, J., Bartl, P., Suder, A., Pajer, P., Kubičková, P., Pavliš, O., Hradilová, M., Vítovská, D., Šanderová, H., Převorovský, M., Hnilicová, J., & Krásný, L. (2019). Ms1 RNA increases the amount of RNA polymerase in *Mycobacterium smegmatis*. *Molecular Microbiology*, *111*(2), 354–372. https://doi.org/10.1111/mmi.14159

Singha, B., Behera, D., Khan, M. Z., Singh, N. K., Sowpati, D. T., Gopal, B., & Nandicoori, V. K. (2023). The unique N-terminal region of *Mycobacterium tuberculosis* sigma factor A plays a dominant role in the essential function of this protein. *Journal of Biological Chemistry*, *299*(3), 102933. https://doi.org/10.1016/j.jbc.2023.102933

Soliveri, J. A., Gomez, J., Bishai, W. R., & Chater, K. F. (2000). Multiple paralogous genes related to the *Streptomyces coelicolor* developmental regulatory gene whiB are present in *Streptomyces* and other *actinomycetes*. *Microbiology*, *146*(2), 333–343. https://doi.org/10.1099/00221287-146-2-333

Song, E., Uhm, H., Munasingha, P. R., Hwang, S., Seo, Y.-S., Kang, J. Y., Kang, C., & Hohng, S. (2022). Rho-dependent transcription termination proceeds via three routes. *Nature Communications*, *13*(1), 1663. https://doi.org/10.1038/s41467-022-29321-5

Spanogiannopoulos, P., Waglechner, N., Koteva, K., & Wright, G. D. (2014). A rifamycin inactivating phosphotransferase family shared by environmental and pathogenic bacteria. *Proceedings of the National Academy of Sciences*, *111*(19), 7102–7107. https://doi.org/10.1073/pnas.1402358111

Srivastava, D. B., Leon, K., Osmundson, J., Garner, A. L., Weiss, L. A., Westblade, L. F., Glickman, M. S., Landick, R., Darst, S. A., Stallings, C. L., & Campbell, E. A. (2013). Structure and function of CarD, an essential mycobacterial transcription factor. *Proceedings of the National Academy of Sciences*, *110*(31), 12619–12624. https://doi.org/10.1073/pnas.1308270110

Stallings, C. L., Stephanou, N. C., Chu, L., Hochschild, A., Nickels, B. E., & Glickman, M. S. (2009). CarD Is an Essential Regulator of rRNA Transcription Required for *Mycobacterium tuberculosis* Persistence. *Cell*, *138*(1), 146–159. https://doi.org/10.1016/j.cell.2009.04.041

Steyn, A. J. C., Collins, D. M., Hondalus, M. K., Jacobs, W. R., Kawakami, R. P., & Bloom, B. R. (2002). *Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect host survival but is dispensable for *in vivo* growth. *Proceedings of the National Academy of Sciences*, *99*(5), 3147–3152. https://doi.org/10.1073/pnas.052705399

* Storz, G., Vogel, J., & Wassarman, K. M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, *43*(6), 880–891. https://doi.org/10.1016/j.molcel.2011.08.022

Surette, M. D., Waglechner, N., Koteva, K., & Wright, G. D. (2022). HelR is a helicase-like protein that protects RNA polymerase from rifamycin antibiotics. *Molecular Cell*, *82*(17), 3151-3165.e9. https://doi.org/10.1016/j.molcel.2022.06.019

Tetsuo, S., Kan, T., & Hideo, T. (1991). Sequence of hrdB, an essential gene encoding sigma-like transcription factor of *Streptomyces coelicolor* A3(2): homology to principal sigma factors. *Gene*, *107*(1), 145–148. https://doi.org/10.1016/0378-1119(91)90308-X

Thomsen, N. D., Lawson, M. R., Witkowsky, L. B., Qu, S., & Berger, J. M. (2016). Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. *Proceedings of the National Academy of Sciences*, *113*(48). https://doi.org/10.1073/pnas.1616745113

* Uptain, S. M., Kane, C. M., & Chamberlin, M. J. (1997). Basic Mechanisms of Transcript Elongation and its Regulation. *Annual Review of Biochemistry*, *66*(1), 117–172. https://doi.org/10.1146/annurev.biochem.66.1.117

* van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., & van Wezel, G. P. (2020). Ecology and genomics of *Actinobacteria*: new concepts for natural product discovery. *Nature Reviews Microbiology*, *18*(10), 546–558. https://doi.org/10.1038/s41579-020-0379-y

Vaňková Hausnerová, V., Marvalová, O., Šiková, M., Shoman, M., Havelková, J., Kambová, M., Janoušková, M., Kumar, D., Halada, P., Schwarz, M., Krásný, L., Hnilicová, J., & Pánek, J. (2022). Ms1 RNA Interacts With the RNA Polymerase Core in *Streptomyces coelicolor* and Was Identified in Majority of *Actinobacteria* Using a Linguistic Gene Synteny Search. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.848536

Waagmeester, A., Thompson, J., & Reyrat, J. M. (2005). Identifying sigma factors in *Mycobacterium smegmatis* by comparative genomic analysis. *Trends in Microbiology*, *13*(11), 505–509. https://doi.org/10.1016/j.tim.2005.08.009

Wang, L., Fan, D., Chen, W., & Terentjev, E. M. (2015). Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces. *Scientific Reports*, *5*(1), 15159. https://doi.org/10.1038/srep15159

Wassarman, K. M., & Storz, G. (2000). 6S RNA Regulates *E. coli* RNA Polymerase Activity. *Cell*, *101*(6), 613–623. https://doi.org/10.1016/S0092-8674(00)80873-9

Wehner, S., Damm, K., Hartmann, R. K., & Marz, M. (2014). Dissemination of 6S RNA among *Bacteria*. *RNA Biology*, *11*(11), 1467–1478. https://doi.org/10.4161/rna.29894

Weiss, L. A., Harrison, P. G., Nickels, B. E., Glickman, M. S., Campbell, E. A., Darst, S. A., & Stallings, C. L. (2012). Interaction of CarD with RNA Polymerase Mediates *Mycobacterium tuberculosis* Viability, Rifampin Resistance, and Pathogenesis. *Journal of Bacteriology*, *194*(20), 5621–5631. https://doi.org/10.1128/JB.00879-12

Wiedermannová, J., Sudzinová, P., Koval', T., Rabatinová, A., Šanderová, H., Ramaniuk, O., Rittich, Š., Dohnálek, J., Fu, Z., Halada, P., Lewis, P., & Krásný, L. (2014). Characterization of HelD, an interacting partner of RNA polymerase from *Bacillus subtilis*. *Nucleic Acids Research*, *42*(8), 5151–5163. https://doi.org/10.1093/nar/gku113

Wildermuth, H., Wehrli, E., & Horne, R. W. (1971). The surface structure of spores and aerial mycelium in *Streptomyces coelicolor*. *Journal of Ultrastructure Research*, *35*(1–2), 168–180. https://doi.org/10.1016/S0022-5320(71)80149-1

Williamson, M. P., Green, J., & Hounslow, A. M. (2017). *M. tuberculosis* [4Fe-4S] protein WhiB1 is a four-helix bundle that forms a NO-sensitive complex with sigmaA and regulates the major virulence factor ESX-1. https://doi.org/10.2210/pdb5oay/pdb

* Wolf, S., Becker, J., Tsuge, Y., Kawaguchi, H., Kondo, A., Marienhagen, J., Bott, M., Wendisch, V. F., & Wittmann, C. (2021). Advances in metabolic engineering of *Corynebacterium glutamicum* to produce high-value active ingredients for food, feed, human health, and well-being. *Essays in Biochemistry*, *65*(2), 197–212. https://doi.org/10.1042/EBC20200134

Zainuddin, Z. F., Kunze, Z. M., & Dale, J. W. (1989). Transformation of *Mycobacterium smegmatis* with *Escherichia coii* plasmids carrying a selectable resistance marker. *Molecular Microbiology*, *3*(1), 29–34. https://doi.org/10.1111/j.1365-2958.1989.tb00100.x

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., & Darst, S. A. (1999). Crystal Structure of *Thermus aquaticus* Core RNA Polymerase at 3.3 A ˚ Resolution DNA template, and finally releases itself and the completed transcript from the DNA when a specific termination signal is encountered. The current view is that the. *Cell*, *98*, 811–824.

Zhu, D. X., Garner, A. L., Galburt, E. A., & Stallings, C. L. (2019). CarD contributes to diverse gene expression outcomes throughout the genome of *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, *116*(27), 13573–13581. https://doi.org/10.1073/pnas.1900176116